

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2004-310607

(43)Date of publication of application : 04.11.2004

(51)Int.Cl.

G06F 17/30
G06F 17/21

(21)Application number : 2003-105656

(71)Applicant : RICOH CO LTD

(22)Date of filing : 09.04.2003

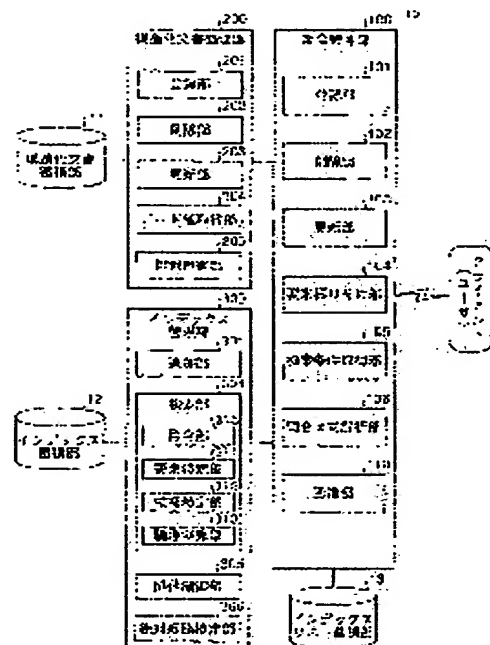
(72)Inventor : TAKAHASHI NOZOMI
KUNIEDA TAKAYUKI
WAKITA YOSHIKI
YAMAGATA JUNICHI
KOYAMA TAKESHI

(54) DOCUMENT MANAGEMENT DEVICE, SYSTEM, METHOD, AND PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a document management device, system, method, and program, with which desired information can be efficiently obtained from a structured document.

SOLUTION: The document management device 10 for managing a document to be accumulated by a document accumulating means 11 by a hierarchical structure is provided with an index generation means 101 for generating an index to specify a position in the hierarchical structure of elements included in the document, and an index management means 300 for performing management by correlating the index generated by the index generation means 101 to the element.



LEGAL STATUS

[Date of request for examination]

30.06.2005

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2004-310607

(P2004-310607A)

(43) 公開日 平成16年11月4日 (2004.11.4)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
G06F 17/30	G06F 17/30 230Z	5B009
G06F 17/21	G06F 17/30 140	5B075
	G06F 17/30 170Z	
	G06F 17/21 501T	
	G06F 17/21 590E	

審査請求 未請求 請求項の数 20 O L (全 35 頁)

(21) 出願番号	特願2003-105656 (P2003-105656)	(71) 出願人	000006747 株式会社リコー 東京都大田区中馬込1丁目3番6号
(22) 出願日	平成15年4月9日 (2003.4.9)	(74) 代理人	100089118 弁理士 酒井 宏明
		(72) 発明者	高橋 望 東京都大田区中馬込1丁目3番6号 株式 会社リコー内
		(72) 発明者	國枝 孝之 東京都大田区中馬込1丁目3番6号 株式 会社リコー内
		(72) 発明者	脇田 由喜 東京都大田区中馬込1丁目3番6号 株式 会社リコー内

最終頁に続く

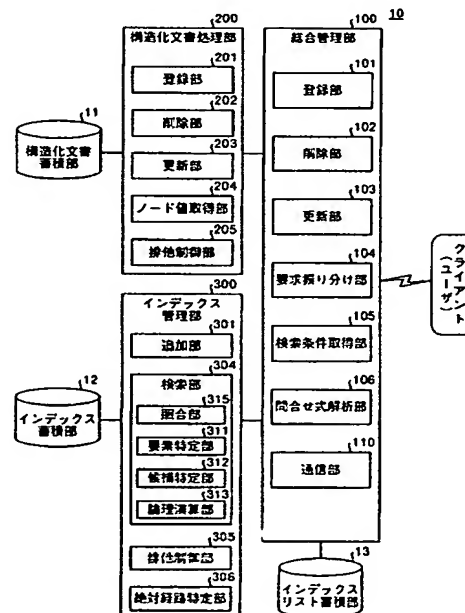
(54) 【発明の名称】 文書管理装置、文書管理システム、文書管理方法、および文書管理プログラム

(57) 【要約】

【課題】 構造化文書から所望の情報を効率的に得ることのできる文書管理装置、文書管理システム、文書管理方法、および文書管理プログラムを提供する。

【解決手段】 文書蓄積手段11に蓄積される文書を階層構造により管理する文書管理装置10であって、前記文書に含まれる要素の前記階層構造における位置を特定するインデックスを生成するインデックス生成手段101と、前記インデックス生成手段101によって生成された前記インデックスと、前記要素とを対応付けて管理するインデックス管理手段300とを備えた。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理装置であって、

前記文書に含まれる要素の前記階層構造における位置を特定するインデックスを生成するインデックス生成手段と、

前記インデックス生成手段によって生成された前記インデックスと、前記要素とを対応付けて管理するインデックス管理手段と

を備えたことを特徴とする文書管理装置。

【請求項 2】

前記インデックスと前記要素とを対応付けて蓄積するインデックス蓄積手段をさらに備え、

前記インデックス管理手段は、前記インデックス蓄積手段に蓄積されている前記インデックスを管理することを特徴とする請求項 1 に記載の文書管理装置。

【請求項 3】

前記要素を検索する検索条件を取得する検索条件取得手段と、

前記検索条件取得手段が取得した前記インデックスと前記インデックス管理手段に管理されている前記インデックスとを照合する照合手段と、

前記照合手段の照合結果に基づいて、前記検索条件に合致する前記要素を特定する要素特定手段と

をさらに備えたことを特徴とする請求項 1 または 2 に記載の文書管理装置。

【請求項 4】

前記インデックス生成手段は、前記要素の前記階層構造における探索経路を示す要素位置特定経路情報と、同一の前記要素位置特定経路情報で特定される複数の要素を識別する経路内要素特定情報とを含む前記インデックスを生成することを特徴とする請求項 1 から 3 のいずれか一項に記載の文書管理装置。

【請求項 5】

前記検索条件取得手段は、前記要素位置特定経路情報および前記経路内要素特定情報のうち少なくとも一方を検索条件として取得し、

前記照合手段は、前記インデックス取得手段が取得した前記検索条件と、前記インデックス管理手段に管理されている前記インデックスとを照合し、

前記要素特定手段は、照合の結果に基づいて、前記検索条件に合致する部分を有する前記インデックスにより、前記検索条件に合致する前記要素を特定することを特徴とする請求項 4 に記載の文書管理装置。

【請求項 6】

前記検索条件取得手段が、複数の検索条件と、各検索条件に対する論理演算子を取得し、前記照合手段が、各検索条件と前記インデックス管理手段に管理されている前記インデックスとをそれぞれ照合した場合に、前記照合の結果に基づいて、各検索条件に合致する前記インデックスをそれぞれ特定する候補特定手段と、

前記候補特定手段が各検索条件に対して特定した複数の前記インデックスに対し、前記論理演算子を用いた論理演算を行う論理演算手段と

をさらに備え、

前記要素特定手段は、当該論理演算の結果得られた前記インデックスにより、検索条件に合致する前記要素を特定することを特徴とする請求項 3 から 5 のいずれか一項に記載の文書管理装置。

【請求項 7】

前記照合手段は、前記インデックスに文字列が含まれる場合に、各検索条件の文字列と前記インデックス管理手段に管理されている前記インデックスに含まれる文字列とを照合し、

前記候補特定手段は、前記文字列の照合の結果に基づいて、各検索条件に合致する前記イ

10

20

30

40

50

ンデックスをそれぞれ特定し、

前記論理演算手段は、特定した複数の前記インデックス間で文字列を照合することにより前記論理演算を行うことを特徴とする請求項 6 に記載の文書管理装置。

【請求項 8】

前記検索条件取得手段は、前記要素特定手段が前回特定した前記要素の前記階層構造における位置から、次に特定すべき要素までの相対経路を取得し、

前記要素特定手段が前回特定した前記要素の前記インデックスと、前記検索条件取得手段が取得した前記相対経路に基づいて、次に特定すべき要素の前記階層構造における絶対経路で示された前記インデックスを特定する絶対経路特定手段をさらに備えたことを特徴とする請求項 3 から 7 のいずれか一項に記載の文書管理装置。

10

【請求項 9】

前記要素特定手段は、前記相対経路特定手段によって特定された前記絶対経路に基づいて、前記検索条件に合致する要素を特定することを特徴とする請求項 8 に記載の文書管理装置。

【請求項 10】

前記要素特定手段が特定した前記要素と、当該要素の前記インデックスとを対応付けて出力する出力手段と

をさらに備えたことを特徴とする請求項 2 から 9 のいずれか一項に記載の文書管理装置。

【請求項 11】

前記インデックス蓄積手段に蓄積されている前記インデックスの更新を行う更新手段をさらに備えたことを特徴とする請求項 1 から 10 のいずれか一項に記載の文書管理装置。

20

【請求項 12】

前記インデックス蓄積手段に蓄積されているインデックスの登録および削除を行う登録削除手段をさらに備えたことを特徴とする請求項 1 から 11 のいずれか一項に記載の文書管理装置。

【請求項 13】

前記更新手段が一の処理を行っている場合に、他の処理を排除する排除制御手段をさらに備えたことを特徴とする請求項 11 に記載の文書管理装置。

【請求項 14】

前記インデックス管理手段は、前記インデックス蓄積手段に蓄積されている前記インデックスの登録、削除、および更新のうち少なくとも一つが行われた場合に、前記インデックスおよび前記要素のうち少なくともいずれか一方を更新することを特徴とする請求項 12 または 13 に記載の文書管理装置。

30

【請求項 15】

文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理装置であって、

前記要素を検索する検索条件を取得する検索条件取得手段と、

前記インデックス生成手段によって生成された前記インデックスと、前記要素とを対応付けて管理するインデックス管理手段と、

前記検索条件取得手段が取得した前記インデックスと前記インデックス管理手段に管理されている前記インデックスとを照合する照合手段と、

40

前記照合手段の照合結果に基づいて、前記検索条件に合致する前記要素を特定する要素特定手段と

を備えたことを特徴とする文書管理装置。

【請求項 16】

文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理システムであって、

前記文書に含まれる要素の前記階層構造における位置を特定するインデックスを生成するインデックス生成手段と、

前記インデックス生成手段によって生成された前記インデックスと前記文書に含まれる要

50

素とを対応付けて蓄積する複数のインデックス蓄積手段と、
前記複数の前記インデックス蓄積手段に蓄積される前記インデックスを管理するインデックス管理手段と
を備えたことを特徴とする文書管理システム。

【請求項 17】

文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理システムであって、

前記文書に含まれる要素の前記階層構造における位置を特定するインデックスを生成するインデックス生成手段と、

前記インデックス生成手段によって生成された前記インデックスと前記文書に含まれる要素とを対応付けて蓄積する複数のインデックス蓄積手段と、

前記複数の前記インデックス蓄積手段に蓄積される前記インデックスをそれぞれ管理する複数のインデックス管理手段と

を備えたことを特徴とする文書管理システム。

【請求項 18】

前記文書を検索する検索条件を取得する検索条件取得手段と、

前記検索条件取得手段が取得した前記インデックスと前記インデックス管理手段に管理されている前記インデックスとを照合する照合手段と、

前記照合手段の照合結果に基づいて、前記検索条件に合致する前記要素を特定する要素特定手段と

をさらに備えたことを特徴とする請求項 16 または 17 に記載の文書管理システム。

【請求項 19】

文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理方法であって、

前記文書の前記階層構造における位置を特定するインデックスを生成するインデックス生成ステップと、

前記インデックス生成ステップにおいて生成された前記インデックスと、前記文書に含まれる要素とを対応付けて、インデックス蓄積手段に蓄積するインデックス蓄積ステップとを有することを特徴とする文書管理方法。

【請求項 20】

文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理方法をコンピュータに実行させる文書管理プログラムであって、

前記文書の前記階層構造における位置を特定するインデックスを生成するインデックス生成ステップと、

前記インデックス生成ステップにおいて生成された前記インデックスと、前記文書に含まれる要素とを対応付けて管理するインデックス管理ステップと

をコンピュータに実行させることを特徴とする文書管理プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、文書蓄積手段に蓄積される文書を階層構造により管理する文書管理装置、文書管理システム、文書管理方法、および文書管理プログラムに関する。

【0002】

【従来の技術】

近年の情報化に伴い電子化文書が増加し、電子化文書を効率的に管理、検索する技術の提供が望まれている。電子化文書を検索する方法としては、論理構造を有する構造化文書によるものが知られている。

【0003】

構造化文書を利用することにより、所定のキーワードを指定することにより、効率的に該当する要素を抽出することができる。

10

20

30

40

50

【 0 0 0 4 】

構造化文書の記述方式としては、W3C (World Wide Web Consortium) で規格化が薦められているXML (extended Markup Language) が知られている。これは、文書への属性情報付加や、文書の階層構造を表現するのに適した記述方式である。このように、属性情報付加や、階層構造を表現するのに適しているという特徴を有することから、新聞や本などの通常の文書だけでなく、出納データや、部品番号データなどの情報を管理するのにも適している。

【 0 0 0 5 】

【 発明が解決しようとする課題 】

しかし、例えば、新聞の全文検索においては、新聞のレイアウト情報は不要である。また、映像データの検索においては、詳細な映像フォーマットに関する情報は不要である。このように、対象となる文書が異なると、検索対象となるデータの種類も異なってくる。このように、構造化文書の対象となる文書は多岐に渡っており、各文書の種類ごとに最適な管理方法も異なってくる。そのため、各文書の種類に適した管理方法の提供が望まれている。また、対象となる文書内の全ての要素が検索対象となるわけではないので、検索対象となる要素のみを効率的に検索することにより、検索などの処理の効率化を図る必要がある。

【 0 0 0 6 】

この発明は、上記に鑑みてなされたもので、所望の情報を効率的に得ることのできる文書管理装置、文書管理システム、文書管理方法、および文書管理プログラムを得ることを目的とする。

【 0 0 0 7 】

【 課題を解決するための手段 】

上記目的を達成するため、請求項1にかかる発明は、文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理装置であって、前記文書に含まれる要素の前記階層構造における位置を特定するインデックスを生成するインデックス生成手段と、前記インデックス生成手段によって生成された前記インデックスと、前記要素とを対応付けて管理するインデックス管理手段とを備えたことを特徴とする。

【 0 0 0 8 】

この請求項1の発明によれば、インデックス管理手段は、文書の階層構造における位置を特定するインデックスと、要素とを対応付けて管理しているので、所定の要素が特定された場合に、当該要素の構造化文書内における位置を認識することができる。このように、特定した要素が構造化文書内のいずれの位置に格納されているのかを把握することができるので、例えば、さらに他の要素を特定する場合に、既に特定している要素の位置を基準とした、他の要素までの相対経路に基づいて他の要素を特定することができる。このように、特定した要素の位置に基づいて、検索処理を行うことにより、処理の効率化を図ることができる。

【 0 0 0 9 】

また、RDBMS (Relation Database Management System) においては構造化文書のスキーマを予め定めておく必要があるが、本発明においては、このようにインデックスと要素とを対応付けているので、構造化文書のスキーマを予め定める必要がない。したがって、文書のスキーマに制限されることなく、多様な種類の文書を管理することができる。

【 0 0 1 0 】

また、請求項2にかかる発明は、請求項1に記載の文書管理装置であって、前記インデックスと前記要素とを対応付けて蓄積するインデックス蓄積手段をさらに備え、前記インデックス管理手段は、前記インデックス蓄積手段に蓄積されている前記インデックスを管理することを特徴とする。

【 0 0 1 1 】

この請求項2の発明によれば、インデックス蓄積手段は、インデックスと要素とを対応付

10

20

30

40

50

けて蓄積しているので、インデックス管理手段は、要素とともにインデックスを特定することができる。また、このインデックスは、構造化文書における要素の位置を示しているため、これにより、要素の位置を把握することができる。このように、特定した要素の位置に基づいて、処理を行うことにより、処理の効率化を図ることができる。

【0012】

また、請求項3にかかる発明は、請求項1または2に記載の文書管理装置であって、前記要素を検索する検索条件を取得する検索条件取得手段と、前記検索条件取得手段が取得した前記インデックスと前記インデックス管理手段に管理されている前記インデックスとを照合する照合手段と、前記照合手段の照合結果に基づいて、前記検索条件に合致する前記要素を特定する要素特定手段とをさらに備えたことを特徴とする。

10

【0013】

この請求項3の発明によれば、インデックスは要素の位置を示しているため、要素の位置を示す検索条件を取得した場合には、検索条件とインデックスとの照合結果に基づいて、検索条件に合致する要素を特定することができる。従って、より効率的に検索対象を特定することができる。

【0014】

また、請求項4にかかる発明は、請求項1から3のいずれか一項に記載の文書管理装置であって、前記インデックス生成手段は、前記要素の前記階層構造における探索経路を示す要素位置特定経路情報と、同一の前記要素位置特定経路情報で特定される複数の要素を識別する経路内要素特定情報とを含む前記インデックスを生成することを特徴とする。

20

【0015】

この請求項4の発明によれば、インデックス生成手段は、階層構造における探索経路を示す要素位置特定経路情報と、同一の要素位置特定経路情報で特定される複数の要素を識別する経路内要素特定情報とを含むインデックスを生成する。このため、同一の経路で特定される要素が複数ある場合、すなわち一の親ノードに対して、複数の子ノードが存在するような構成であっても、複数の要素それぞれを識別することができる。従って、すべての要素を正確に識別することができる。

【0016】

また、請求項5にかかる発明は、請求項4に記載の文書管理装置であって、前記検索条件取得手段は、前記要素位置特定経路情報および前記経路内要素特定情報のうち少なくとも一方を検索条件として取得し、前記照合手段は、前記インデックス取得手段が取得した前記検索条件と、前記インデックス管理手段に管理されている前記インデックスとを照合し、前記要素特定手段は、照合の結果に基づいて、前記検索条件に合致する部分を有する前記インデックスにより、前記検索条件に合致する前記要素を特定することを特徴とする。

30

【0017】

この請求項5の発明によれば、検索条件が要素位置特定経路情報および経路内要素特定情報のうち少なくとも一方を含む場合には、当該検索条件とインデックスとを照合することによって、検索条件に合致する要素を特定することができる。従って、より効率的に検索対象を特定することができる。

【0018】

また、請求項6にかかる発明は、請求項3から5のいずれか一項に記載の文書管理装置であって、前記検索条件取得手段が、複数の検索条件と、各検索条件に対する論理演算子を取得し、前記照合手段が、各検索条件と前記インデックス管理手段に管理されている前記インデックスとをそれぞれ照合した場合に、前記照合の結果に基づいて、各検索条件に合致する前記インデックスをそれぞれ特定する候補特定手段と、前記候補特定手段が各検索条件に対して特定した複数の前記インデックスに対し、前記論理演算子を用いた論理演算を行う論理演算手段とをさらに備え、前記要素特定手段は、当該論理演算の結果得られた前記インデックスにより、検索条件に合致する前記要素を特定することを特徴とする。

40

【0019】

この請求項6の発明によれば、複数の検索条件と、各検索条件に対する論理演算子を取得

50

した場合には、まず、各検索条件とインデックスとの照合結果に基づいて、各検索条件に合致するインデックスを特定する。そして、特定したインデックスに対して論理演算子を用いた論理演算を行うことにより、検索対象を示すインデックスを特定することができる。このように、インデックスと検索対象との照合により検索対象を特定することができるので、処理効率を向上させることができる。

【0020】

また、請求項7にかかる発明は、請求項6に記載の文書管理装置であって、前記照合手段は、前記インデックスに文字列が含まれる場合に、各検索条件の文字列と前記インデックス管理手段に管理されている前記インデックスに含まれる文字列とを照合し、前記候補特定手段は、前記文字列の照合の結果に基づいて、各検索条件に合致する前記インデックスをそれぞれ特定し、前記論理演算手段は、特定した複数の前記インデックス間で文字列を照合することにより前記論理演算を行うことを特徴とする。

10

【0021】

この請求項7の発明によれば、インデックスに文字列が含まれている場合には、検索条件とインデックスに含まれている文字列の各文字を照合することにより、実質的に、論理演算を行うことができる。このように、文字列照合の結果を利用することにより、検索処理の効率を向上させることができる。

【0022】

また、請求項8にかかる発明は、請求項3から7のいずれか一項に記載の文書管理装置であって、前記検索条件取得手段は、前記要素特定手段が前回特定した前記要素の前記階層構造における位置から、次に特定すべき要素までの相対経路を取得し、前記要素特定手段は、前回特定した前記要素の前記インデックスと、前記検索条件取得手段が取得した前記相対経路に基づいて、次に特定すべき要素の前記階層構造における絶対経路を特定する絶対経路特定手段をさらに備えたことを特徴とする。

20

【0023】

この請求項8の発明によれば、絶対経路特定手段は、前回特定したインデックスと、検索条件取得手段が取得した相対経路に基づいて、次に特定すべき要素の絶対経路、すなわちインデックスを特定することができるので、検索条件を解析して構造化文書の根から検索を開始する場合に比べて、検索処理の効率を向上させることができる。

【0024】

また、請求項9にかかる発明は、請求項8に記載の文書管理装置であって、前記要素特定手段は、前記絶対経路特定手段によって特定された前記絶対経路に基づいて、前記検索条件に合致する要素を特定することを特徴とする。

30

【0025】

この請求項9の発明によれば、絶対経路特定手段によって特定された絶対経路に基づいて、検索条件に合致する要素を特定することができる。すなわち、絶対経路を取得しなくとも、相対経路から特定した絶対経路に基づいて、要素を特定することができる。

【0026】

また、請求項10にかかる発明は、請求項2から9のいずれか一項に記載の文書管理装置であって前記要素特定手段が特定した前記要素と、当該要素の前記インデックスとを対応付けて出力する出力手段とをさらに備えたことを特徴とする。

40

【0027】

この請求項10の発明によれば、検索条件に合致するとして抽出した要素と、当該要素のインデックスとを対応付けて出力するので、ユーザは、要素が構造化文書のいずれの位置に格納されているのかを把握することができる。

【0028】

また、請求項11にかかる発明は、請求項1から10のいずれか一項に記載の文書管理装置であって、前記インデックス蓄積手段に蓄積されている前記インデックスの更新を行う更新手段をさらに備えたことを特徴とする。

【0029】

50

この請求項 1 1 の発明によれば、インデックス蓄積手段に蓄積されているインデックスを更新する更新手段をさらに備えているので、インデックスを自由に更新することができる。

【 0 0 3 0 】

また、請求項 1 2 にかかる発明は、請求項 1 から 1 1 のいずれか一項に記載の文書管理装置であって、前記インデックス蓄積手段に蓄積されているインデックスの登録および削除を行う登録削除手段をさらに備えたことを特徴とする。

【 0 0 3 1 】

この請求項 1 2 の発明によれば、インデックス蓄積手段に蓄積されているインデックスの登録および削除を行う登録削除手段をさらに備えているので、インデックスを自由に登録および削除することができる。

10

【 0 0 3 2 】

また、請求項 1 3 にかかる発明は、請求項 1 1 に記載の文書管理装置であって、前記更新手段が一の処理を行っている場合に、他の処理を排除する排除制御手段をさらに備えたことを特徴とする。

【 0 0 3 3 】

この請求項 1 3 の発明によれば、更新手段が一の処理を行っているときに、排他制御手段は、他の処理を排除するので、複数の処理が同時に行われることにより、不適切な更新が行われるのを避けることができる。

【 0 0 3 4 】

また、請求項 1 4 にかかる発明は、請求項 1 2 または 1 3 に記載の文書管理装置であって、前記インデックス管理手段は、前記インデックス蓄積手段に蓄積されている前記インデックスの登録、削除、および更新のうち少なくとも一つが行われた場合に、前記インデックスおよび前記要素のうち少なくともいずれか一方を更新することを特徴とする。

20

【 0 0 3 5 】

この請求項 1 4 の発明によれば、インデックス管理手段は、インデックス蓄積手段に蓄積されているインデックスの登録、削除、および更新のうち少なくとも一つが行われた場合には、インデックス蓄積手段に蓄積されているインデックスのうち、当該処理により更新の必要が生じたデータを更新することができる。従って、一の処理により、他のデータとの整合性がとれなくなるのを回避することができる。

30

【 0 0 3 6 】

また、請求項 1 5 にかかる発明は、文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理装置であって、前記要素を検索する検索条件を取得する検索条件取得手段と、前記インデックス生成手段によって生成された前記インデックスと、前記要素とを対応付けて管理するインデックス管理手段と、前記検索条件取得手段が取得した前記インデックスと前記インデックス管理手段に管理されている前記インデックスとを照合する照合手段と、前記照合手段の照合結果に基づいて、前記検索条件に合致する前記要素を特定する要素特定手段とを備えたことを特徴とする。

【 0 0 3 7 】

この請求項 1 5 の発明によれば、インデックスは要素の位置を示しているので、要素の位置を示す検索条件を取得した場合には、検索条件とインデックスとの照合結果に基づいて、検索条件に合致する要素を特定することができる。従って、より効率的に検索対象を特定することができる。

40

【 0 0 3 8 】

また、請求項 1 6 にかかる発明は、文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理システムであって、前記文書に含まれる要素の前記階層構造における位置を特定するインデックスを生成するインデックス生成手段と、前記インデックス生成手段によって生成された前記インデックスと前記文書に含まれる要素とを対応付けて蓄積する複数のインデックス蓄積手段と、前記複数の前記インデックス蓄積手段に蓄積される前記インデックスを管理するインデックス管理手段とを備えたことを特徴とする。

50

【0039】

この請求項16の発明によれば、複数のインデックス蓄積手段を備えているので、処理の効率化を図ることができる。

【0040】

また、請求項17にかかる発明は、文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理システムであって、前記文書に含まれる要素の前記階層構造における位置を特定するインデックスを生成するインデックス生成手段と、前記インデックス生成手段によって生成された前記インデックスと前記文書に含まれる要素とを対応付けて蓄積する複数のインデックス蓄積手段と、前記複数の前記インデックス蓄積手段に蓄積される前記インデックスをそれぞれ管理する複数のインデックス管理手段とを備えたことを特徴とする。

10

【0041】

この請求項17の発明によれば、複数のインデックス蓄積手段と、各インデックス蓄積手段を管理する複数のインデックス管理手段とを備えているので、分散処理が可能となり、処理の効率化を図ることができる。

【0042】

また、請求項18にかかる発明は、請求項16または17に記載の文書管理システムであって、前記文書を検索する検索条件を取得する検索条件取得手段と、前記検索条件取得手段が取得した前記インデックスと前記インデックス管理手段に管理されている前記インデックスとを照合する照合手段と、前記照合手段の照合結果に基づいて、前記検索条件に合致する前記要素を特定する要素特定手段とをさらに備えたことを特徴とする。

20

【0043】

この請求項18の発明によれば、インデックスは要素の位置を示しているので、要素の位置を示す検索条件を取得した場合には、検索条件とインデックスとの照合結果に基づいて、検索条件に合致する要素を特定することができる。従って、より効率的に検索対象を特定することができる。

【0044】

また、請求項19にかかる発明は、文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理方法であって、前記文書の前記階層構造における位置を特定するインデックスを生成するインデックス生成ステップと、前記インデックス生成ステップにおいて生成された前記インデックスと、前記文書に含まれる要素とを対応付けて、インデックス蓄積手段に蓄積するインデックス蓄積ステップとを有することを特徴とする。

30

【0045】

この請求項19の発明によれば、特定した要素が構造化文書内のいずれの位置に格納されているのかを把握することができるので、例えば、さらに他の要素を特定する場合に、既に特定している要素の位置を基準とした、他の要素までの相対経路に基づいて他の要素を特定することができる。このように、特定した要素の位置に基づいて、検索処理を行うことにより、処理の効率化を図ることができる。

【0046】

また、請求項20にかかる発明は、文書蓄積手段に蓄積される、階層構造により表現された文書を管理する文書管理方法をコンピュータに実行させる文書管理プログラムであって、前記文書の前記階層構造における位置を特定するインデックスを生成するインデックス生成ステップと、前記インデックス生成ステップにおいて生成された前記インデックスと、前記文書に含まれる要素とを対応付けて管理するインデックス管理ステップとをコンピュータに実行させることを特徴とする。

40

【0047】

この請求項20の発明によれば、特定した要素が構造化文書内のいずれの位置に格納されているのかを把握することができるので、例えば、さらに他の要素を特定する場合に、既に特定している要素の位置を基準とした、他の要素までの相対経路に基づいて他の要素を特定することができる。このように、特定した要素の位置に基づいて、検索処理を行うこ

50

とにより、処理の効率化を図ることができる。

【 0 0 4 8 】

【 発 明 の 実 施 の 形 態 】

以下に添付図面を参照して、この発明にかかる文書管理装置および文書管理システムの好適な実施の形態を詳細に説明する。

【 0 0 4 9 】

（実施の形態 1）

図 1 は、この発明の実施の形態 1 である文書管理装置 1 0 の構成を示すブロック図である。文書管理装置 1 0 は、構造化文書蓄積部 1 1 と、構造化文書蓄積部 1 1 を制御する構造化文書管理部 2 0 0 と、インデックス蓄積部 1 2 と、インデックス蓄積部 1 2 を制御するインデックス処理部 3 0 0 と、インデックスリスト蓄積部 1 3 と、インデックスリスト蓄積部 1 3 を制御し、また、文書管理装置 1 0 全体を制御する総管理部 1 0 0 とを備えている。

10

【 0 0 5 0 】

構造化文書蓄積部 1 1 は、構造化文書と、当該構造化文書を識別する構造化文書 I D とを対応付けて蓄積する。図 2 を参照しつつ、構造化文書蓄積部 1 1 における構造化文書のデータ構成について説明する。構造化文書は、階層構造を有しており、文書に含まれる各要素が、各階層に分けて管理されている。

【 0 0 5 1 】

例えば、図 2 に示される要素「" タイトル" 」 4 1 0 は、ノード 4 0 0 から 3 段階下位のノード 4 0 3 に格納されている。すなわち、要素 4 1 0 を特定するための絶対経路は、ノード 4 0 0、ノード 4 0 1、ノード 4 0 2、およびノード 4 0 3 である。また、X P a t h の記述方式によれば、「A u d i o V i s u a l S e g m e n t [1] / C r e a t i o n I n f o r m a t i o n / T i t l e」と表現される。

20

【 0 0 5 2 】

ここで、「A u d i o V i s u a l S e g m e n t [1]」中の [1] は、「A u d i o V i s u a l S e g m e n t」と同一階層の複数のノードを識別するための識別番号である。図 2 に示す階層構造においては、「A u d i o V i s u a l S e g m e n t」 4 0 5、4 1 6 は、それぞれ「A u d i o V i s u a l S e g m e n t [1]」および「A u d i o V i s u a l S e g m e n t [2]」と表される。

30

【 0 0 5 3 】

また、要素「" キーワード" 」 4 1 1 と要素「" 特許 1" 」 4 1 2 の絶対経路は、いずれもノード 4 0 0、ノード 4 0 4、ノード 4 0 5、およびノード 4 0 6、ノード 4 0 7 である。このように、絶対経路が等しい場合には、さらに、経路内 I D により識別される。ここで、絶対経路とは、階層構造における途中の枝であるノードからの相対的な経路ではなく、根であるノードを基準とした絶対的な経路のことである。また、経路内 I D とは、同一の親ノードに対する複数の子ノードを識別するための情報である。

【 0 0 5 4 】

このように、各要素は、絶対経路および経路内 I D により特定される。本実施の形態におけるインデックスは、この絶対経路を識別する絶対経路 I D と経路内 I D で構成されている。ここで、絶対経路 I D は、本発明における要素位置特定経路情報に相当する。また、経路内 I D は、本発明における経路内要素特定情報に相当する。

40

【 0 0 5 5 】

なお、本実施の形態においては、構造化文書は、ファイルとして管理されており、構造化文書 I D はファイル名である。

【 0 0 5 6 】

他の例としては、構造化文書は、メモリに記憶されていてもよい。また、構造化文書 I D は、構造化文書を一意に識別できる情報であればよく、ファイルパスであってもよい。また、他の例としては、U R L であってもよい。

【 0 0 5 7 】

50

また、構造化文書蓄積部 1 1 の実現方式は、実施の形態に限定されるものではなく、他の例としては、データベースで実現されてもよい。

【0058】

構造化文書管理部 2 0 0 は、構造化文書蓄積部 1 1 に構造化文書を新規に登録する登録部 2 0 1 と、構造化文書蓄積部 1 1 に既に蓄積されている構造化文書を削除する削除部 2 0 2 と、構造化文書蓄積部 1 1 に蓄積されている構造化文書の内容等を更新する更新部 2 0 3 と、構造化文書蓄積部 1 1 に蓄積されている構造化文書から所定のノード値を取得するノード値取得部 2 0 4 と、構造化文書蓄積部 1 1 に対して一の処理が行われているときに、他の処理を排除する排他制御部 2 0 5 とを有する。

【0059】

インデックスリスト蓄積部 1 3 は、構造化文書蓄積部 1 1 に蓄積されている構造化文書を構成する各ノードのリスト構造である、インデックスリストを蓄積する。

【0060】

総管理部 1 0 0 は、インデックスリスト蓄積部 1 3 に新たにインデックスリストに登録する登録部 1 0 1 と、インデックスリスト蓄積部 1 3 に蓄積されているインデックスリストを削除する削除部 1 0 2 と、インデックスリスト蓄積部 1 3 に蓄積されているインデックスリストを更新する更新部 1 0 3 と、外部からの情報を取得する通信部 1 1 0 と、通信部 1 1 0 が取得した情報を振り分ける要求振り分け部 1 0 4 と、要求振り分け部 1 0 4 から検索条件を取得し、また、検索条件に対する検索結果を構造化文書管理部 2 0 0、インデックス処理部 3 0 0、またはインデックスリスト蓄積部 1 3 から取得する検索条件、検索条件取得部 1 0 5 と、検索条件取得部 1 0 5 から検索条件を受け取り、検索条件を解析する検索条件解析部 1 0 6 とを有する。ここで、登録部 1 0 1 は、本発明のインデックス生成手段を構成する。

【0061】

インデックス蓄積部 1 2 は、構造化文書蓄積部 1 1 に蓄積されている構造化文書内の所定のノードを特定するために利用されるインデックスを、対応する構造化文書を識別する構造化文書 ID に対応付けて蓄積する。

【0062】

インデックスは、第 1 値と第 2 値のリスト構造で構成されている。ここで、第 1 値は、インデックスにより特定される要素である。また、第 2 値は、インデックスへのアクセス手段の識別子、すなわちインデックス ID である。本実施の形態においては、インデックス ID は、絶対経路 ID、および経路内 ID を含んでいる。ここで、絶対経路 ID は、単一の要素名、または連続する複数の要素名で構成されている。同一の経路内に複数のインデックスが存在する場合、すなわち、兄弟インデックスが存在する場合には、さらに経路内 ID を含んでいる。なお、二値はユーザにより定義される。

【0063】

より、具体的には、例えば、図 2 における要素「キーワード」4 1 1 のインデックスにおける第 1 値は、キーワードであり、第 2 値は、「AudioVisualSegment / Temporal Decomposition / AudioVisualSegment / TextAnnotation FreeTextAnnotation」である。

【0064】

また、要素名が「FreeTextAnnotation」である場合、第 1 値は、「FreeTextAnnotation」である。また、あるノードの要素名が「TextAnnotation」であり、このノードの子ノード要素名が「FreeTextAnnotation」である場合には、当該子ノードの第 1 値は、「TextAnnotation / FreeTextAnnotation」である。

【0065】

また、要素名「TextAnnotation」で属性名「type」の値が「headline」であるノードの子ノードの要素名が「FreeTextAnnotation

10

20

30

40

50

」である場合、当該子ノードの第1値は、「Text Annotation [@ type = " headline"] / FreeText Annotation」である。

また、それぞれのインデックスの第2値は、それぞれのインデックスへのアクセス手段である。具体的には、第2値は、例えば、「Text DB : FreeText Annotation」であってもよい。また、他の例としては、例えば、URLであってもよい。

【0066】

インデックス処理部300は、インデックス蓄積部12にインデックスを追加する追加部301と、インデックス蓄積部12に蓄積されているインデックスを削除する削除部302と、インデックス蓄積部12に蓄積されているインデックスの内容等を更新する更新部303と、インデックス蓄積部12から所定のインデックスを検索する検索部304と、インデックス蓄積部12に対して一の処理が行われているときに、他の処理を排除する排他制御部305と、要素特定手部312が前回特定した要素のインデックスと、検索条件取得105が取得した相対経路に基づいて、次に特定すべき要素の階層構造における絶対経路で示されたインデックスを特定する絶対経路特定部306とを備えている。

【0067】

また、検索部304は、検索条件と、インデックス蓄積部12に蓄積されているインデックスとの文字列照合を行う照合部315と、照合結果に基づいて、検索条件に合致する要素を特定する要素特定部311と、検索条件に合致する要素の候補のインデックスを特定する候補特定部312と、検索条件に含まれる論理演算子による論理演算を行う論理演算部314とを有している。なお、ここで、論理演算部314は、文字列の比較により、結果として論理演算の結果を取得する。

【0068】

ここで、通信部110は、本発明の出力手段を構成する。また、登録部101と削除部102は、本発明の登録削除手段を構成する。

【0069】

図3は、MPEG-7 (Multimedia Content Description Interface) で記述した構造化文書を示している。図3に示す構造化文書は、図2に示す構造化文書と同一である。

【0070】

ここで、MPEG-7とは、動画や音声などのマルチメディアコンテンツを有効に検索するための記述方法の標準技術である。具体的には、登場人物や映像ファイルフォーマットなどのメタデータを構造化文書 (XML) で記述する方式である。

【0071】

MPEG-7では、対象となる映像データに関する多くの情報を記述することができる。しかし、MPEG-7を実際に映像検索システムに利用する場合、詳細な映像フォーマットの情報は検索対象とはならない場合がある。一方、映像フォーマット変換システムでは、詳細な映像フォーマットの情報のみが必要である。また、新聞のレイアウト情報は、新聞の全文検索データベースには不要である。

【0072】

さらに、構造化文書は構造を持った情報を記述するための文書であるため、帳票データや顧客データのようにその構造化文書の全情報が検索対象となるとは限らず、検索対象となるのは、構造化文書内の一部分のみである場合もある。このようにインデックスを付与すべき対象となる情報は、アプリケーション毎に、またはデータ毎に異なっている。

【0073】

利用されない情報に対してインデックスを付与し、管理することは、処理効率を低下させる要因となる。このように、多様なデータに対して、それぞれのデータ特性にあったインデックスを付与することによって、より効率的に構造化文書を管理することができる。なお、利用されない情報には、XMLにおけるタグ情報も含まれる。

【0074】

10

20

30

40

50

次に、動画データの構造化文書について説明する。動画データでは、開始時間と、時間長により所定のフレームを特定する。ただし、その特定方法はDTD (Document Type Definition) やXML Schemeからは一意に決められない場合がある。なお、DTDとは、タグの属性などを定義する言語であり、XML Schemeとは、文書の構造を定義する言語である。

【0075】

図4は、動画データにおける一部のデータを指定する記述を示している。一部のデータとは、例えば、動画データの開始時間から30秒後に再生されるデータのことである。一部のデータは、開始時間と時間長により指定される。図4(A)、(B)、および(C)は、いずれも開始時間30秒後、時間長20秒を表現している。いずれも、2行目で開始時間を指定している。また、いずれも3行目で時間長を示している。

【0076】

図4(A)は、「". . . / . . . / MediaTime [1]"」の時間点PTの30秒後、すなわち「PT30S」という直接的な表現で表している。なお、ここで、「. . .」であれば軸方向は“上り”で、その数は1であることを示している。図4(B)は、「". . . / . . . / MediaTime [1]"」の時間点PTの10秒単位、すなわち「timeUnit="PT10S"」の3個分後、すなわちボディ部に「3」と表現している。また、図4(C)は、30秒単位「timeUnit="PT30S"」を1個分と表現している。

【0077】

このように複数の記述を許容すると、例えば「". . . / . . . / MediaTime [1]"」で記述される開始時間から10秒後のMediaTimeの全てを探すという検索条件を受けた場合には、開始時間を表現する二種類の要素名「MediaRelTimePoint」と「MediaRelIncrTimePoint」のインデックスを参照し、「MediaRelTimePoint」の値を解釈し検索条件と比較し、「MediaRelIncrTimePoint」のボディ部と属性部を解釈し、検索条件と比較するという処理を行わなければならない。2種類の場合には、さほど処理量の増加は問題とならない。しかし、100種類の表現形式で同一の内容を記述した場合には、処理量が膨大となり、処理効率が低下するという問題が生じる。

【0078】

また、「". . . / . . . / MediaTime [1]"」で記述される終了時間から40秒後のMediaTimeの全てを探すという検索条件を受けた場合には、図4(A)の場合には、二種類の要素「MediaRelTimePoint」および「MediaIncrDuration」を参照して、それぞれ終了時刻を計算する必要あり効率が悪い。

【0079】

図5は、本実施の形態における動画データの構造化文書510を示している。構造化文書510内の行512, 513, 514は、いずれも動画データ中の所定の時間を指定する記述である。すなわち、動画データ中の一部のデータを指定する記述である。時間を指定する4行のうち2行目512a, 513a, 514aは、いずれも開始時間を示している。また、3行目512b, 513b, 514bは、いずれも開始時間からの時間長を示している。

【0080】

このように、本実施の形態においては、時間を指定する情報は、統一された形式で記述されており、かつ、他の情報と同様に絶対経路IDとして記述されている。このため、図4を参照しつつ説明したように、同一の情報を指定する方法が複数存在することにより、処理効率が低下するという問題を解消することができる。

【0081】

さらに、本実施の形態における構造化文書のインデックスは、W3Cで規格化されているXPath (XML Path language) の方式で表現されている。構造化文書の記述方式の代表的なものに、W3Cで規格化が進められているXMLがある。XPa

10

20

30

40

50

thとは、XML文書の記述を特定する方法である。

【0082】

以下、図6を参照しつつ、XPath式による記述について説明する。図6は、XPath式により示された絶対経路IDを示している。図6(A)は、XML方式で示した絶対経路IDを示している。図6(B)は、図6(A)に示す絶対経路IDを、XPath式で表している。

【0083】

このように、絶対経路IDをXPath式で表すことにより、同じ親ノードに属するノードは、数字部分のみが異なる文字列で表現されるので、単純な文字列の比較により各ノードを識別することができる。

【0084】

図7は、インデックス蓄積部12に格納されているインデックスのデータ構造を示している。インデックス520は、第1値と第2値を含むリスト構造で構成されている。第1値は、絶対経路ID522である。また、第2値は、絶対経路ID522により特定される値524である。

【0085】

ここで、絶対経路IDにより特定される値は、構造化文書に含まれる情報であるが、他の例としては、予め計算された値であってもよい。ここで、予め計算された値とは、例えば、図4を参照しつつ説明した動画データの一部のデータのことである。すなわち、開始時間と時間長により特定される一部のデータのことである。

【0086】

なお、インデックス蓄積部12は、RDB(Relation Data Base)である必要はない。

【0087】

従来は、検索エンジンとしてRDBMSの利用を想定しているため、構造化文書中に例えば複雑な計算を要する画像特徴量が記述されている場合には、その画像特徴量に対する問合せも他の文字列の問合せと同時にを行う必要があるアプリケーションを考えた場合、複雑な画像特徴量問い合わせ処理も文字列問い合わせ用RDBMSで行わせるか、または画像特徴量検索エンジンにRDBMSとしてのインターフェイスを持たせ、問合せ式を解析し検索最適化モジュールで文字列問い合わせ用RDBMSと画像特徴量用DBに割り振らなければならず、処理が煩雑であった。

【0088】

これに対して、本実施の形態におけるインデックス蓄積部12に格納されるインデックスは、図7に示すように、第1値および第2値の2つの情報を有するという簡単な形式で構成されている。従って、検索条件を受け付けたときに、対応するインデックスを返す機能さえあれば、特にRDBMSに限定する必要はない。このため、あるキーに対して値を返すといういわゆる検索エンジンを簡単に組み込むことができる。

【0089】

図8は、文書管理装置10の新規文書登録処理を示すフローチャートである。新規文書登録処理とは、構造化文書蓄積部11に新たに文書を蓄積する処理である。

【0090】

通信部110は、クライアントから文書を登録する旨の登録要求を受信し(ステップS100)、要求振り分け部104に送る。要求振り分け部104が受け取った要求が登録要求である場合には(ステップS101, Yes)、登録部101が呼び出され、登録処理が開始する。

【0091】

まず、登録部101は、構造化文書管理部200の登録部201を介して、構造化文書蓄積部11に対象となる構造化文書を登録する(ステップS102)。なお、本実施の形態においては、通信部110が、外部から、すなわち登録要求の送信元であるクライアントから構造化文書の全文を取得する。他の例としては、構造化文書を参照するためのファイ

10

20

30

40

50

ルパスを取得する。この場合には、取得したファイルパスに基づいて、構造化文書を取得する。なお、構造化文書および構造化文書のファイルパスのいずれも取得できない場合には、通信部 110 は、クライアントに対して、エラーを返す。

【0092】

次に、登録部 101 は、構造化文書蓄積部 11 に登録した文書に対する構造化文書 ID を取得する（ステップ S103）。本実施の形態においては、登録部 101 は、構造化文書管理部 200 を介して、構造化文書のファイル名を構造化文書 ID として構造化文書蓄積部 11 から抽出する。

【0093】

次に、登録部 101 は、通信部 110 を介して、構造化文書蓄積部 11 に蓄積されている文書のインデックスリストを取得する（ステップ S104）。次に、インデックス処理部 300 の排他制御部 305 は、インデックス蓄積部 12 に対して排他制御を行う（ステップ S106）。具体的には、インデックス蓄積部 12 は、一の要求に応じて処理を開始すると、当該処理が終了するまで、他の処理を受け付けない。

【0094】

次に、インデックス処理部 300 の追加部 301 が呼び出される。そして、追加部 301 は、当該構造化文書のインデックスをインデックス蓄積部 12 に追加する（ステップ S110）。次に、インデックス処理部 300 の排他制御部 305 は、全てのインデックスに対する排他制御を解除する（ステップ S112）。そして、総合管理部 100 の通信部 110 は、クライアントに対して、登録処理が完了した旨を通知する（ステップ S114）。以上で、登録処理が完了する。

【0095】

図 9 は、図 8 におけるインデックス追加処理（ステップ S110）における詳細な処理を示すフローチャートである。

【0096】

インデックス追加処理では、まず、総合管理部 100 の登録部 101 は、現在登録処理の対象となっている構造化文書が複数ある場合には、そのうちの一の構造化文書に対して、XPath を指定する（ステップ S200）。そして、XPath に対する返答として、インデックスリスト蓄積部 13 から該当する XML ノードの集合を取得する（ステップ S201）。

【0097】

具体的には、構造化文書と、第 1 値を指定して、インデックスへの追加を要求する。ここで、第 1 値は、例えば、「FreeTextAnnotation」など、登録するノードの要素名である。

【0098】

次に、未処理のノードがある場合には（ステップ S202, Yes）、未処理のノードに対して絶対経路 ID を定義する（ステップ S204）。次に、兄弟ノードが存在する場合には、経路内 ID を定義する（ステップ S206）。すなわち、既に登録されているノードと今回登録するノードを兄弟とし、両者を識別する経路内 ID を定義する。

【0099】

具体的には、ノードが既に 1 つ存在していた場合には、既に存在していたノードを今回登録するノードの兄ノードとし、当該兄ノードの同一経路内 ID を「1」と定義する。さらに、今回登録するノードの同一経路内 ID を「2」と定義する。

【0100】

さらに、当該ノードに対する上位の階層がある場合には（ステップ S208, Yes）、上位の階層のノードの経路内 ID を定義する（ステップ S206）。以上、ステップ S206 およびステップ S208 の処理を繰り返し、今回登録するノードに対する上位の全ての階層におけるノードに対して経路内 ID を定義すると（ステップ S208, No）、今回登録するノードの絶対経路 ID を決定する（ステップ S210）。そして、今回登録するノードに対して割り当てられた絶対経路 ID をインデックス蓄積部 12 に追加する（ステ

10

20

30

40

50

ップS 2 1 2)。さらに、インデックスリスト蓄積部 1 3に今回登録するノードのインデックスリストを追加する(ステップS 2 1 4)。

【0101】

以上、ステップS 2 0 4からステップS 2 1 4の処理を、登録すべきノードに対して繰り返し(ステップS 2 0 2)、全てのノードに対して処理が完了すると、ステップS 1 1 0の処理が完了し、図8に示すステップS 1 1 2へ進む。

【0102】

本実施の形態においては、構造的に連続した要素、すなわち、階層における深さ方向に対応したインデックスを作成する。従って、図10に示す構造化文書に対してインデックスを追加する場合には、「Location」の子ノードが「Name」である場合のみインデックスを生成すると指定することにより、「Pattern」を含むノードを除外して、「東京」を含むノードのみをインデックス追加の対象として選択することができる。

10

【0103】

次に、図11を参照しつつ、既に登録されているインデックスを削除する場合の処理について説明する。

【0104】

通信部110がクライアントから文書を削除する旨の削除要求を受信し(ステップS 3 0 0)、要求振り分け部104に送る。要求振り分け部104が受け取った要求が削除要求である場合に(ステップS 3 0 1, Yes)、削除部102が呼び出され、処理が開始する。なお、削除要求には、削除の対象となる構造化文書の構造化文書IDが含まれている。削除要求に構造化文書IDが含まれていない場合には、クライアントに対してエラーを返す。

20

【0105】

まず、削除部102は、インデックスリスト蓄積部13からインデックスリストを取得する(ステップS 3 0 2)。次に、インデックス処理部300の排他制御部305に対して、削除部102が取得したインデックスリストに含まれているインデックスの排他制御を要求する(ステップS 3 0 3)。

【0106】

次に、インデックスリストの第2値、すなわち絶対経路IDを参照し、絶対経路IDに基づいてインデックス処理部300の削除部302に削除すべき構造化文書IDを指定する(ステップS 3 0 4)。削除部302は、指定された構造化文書IDにより特定されるインデックスをインデックス蓄積部12から削除する(ステップS 3 0 6)。具体的には、まず、構造化文書IDが指定されて、削除部302が呼び出される。そして、削除条件として、インデックスの第2値の先頭文字列が「document(「構造化文書」)」であるインデックスを指定して、インデックス蓄積部12に対して削除を要求する。

30

【0107】

以上、ステップS 3 0 4およびステップS 3 0 6の処理を指定された全てのインデックスに対して行う(ステップS 3 0 8)。

【0108】

インデックス蓄積部12から全てのインデックスを削除した後(ステップS 3 0 8, Yes)、構造化文書管理部200の削除部202は、指定された構造化文書IDにより特定される構造化文書を構造化文書蓄積部11から削除する(ステップS 3 1 0)。具体的には、まず構造化文書IDが指定されて、削除部202が呼び出される。そして、構造化文書IDにより特定される構造化文書を削除対象として、構造化文書蓄積部11に削除を要求する。

40

【0109】

次に、インデックス処理部300の排他制御部305は、全インデックスに対する排他制御を解除し(ステップS 3 1 2)、さらに、総管理部100の通信部110は、削除処理が完了した旨をクライアントに通知する(ステップS 3 1 4)。以上で、削除処理が終了する。

50

【0110】

以上のように、インデックスの第2値に対する条件が指定された削除の要求に対して、該当する値がリスト構造内に含まれている場合には、このインデックスを全て削除することができる。

【0111】

他の例としては、インデックスの第1値に対する条件が指定された削除の要求に対して、該当するインデックスを全て削除してもよい。また、インデックスの第1値に対する条件と第2値に対する条件とがそれぞれ指定された削除要求に対して、該当するインデックスを全て削除してもよい。

【0112】

次に、図12を参照しつつ、更新処理について説明する。更新処理とは、構造化文書蓄積部11に格納された構造化文書に対するインデックスの内容を更新する処理である。

【0113】

通信部110がクライアントからインデックスの内容を更新する旨の更新要求を受信し、要求振り分け部104に送る。要求振り分け部104は、更新要求を受け取ると、更新部103を呼び出し、処理を開始する。

【0114】

まず、更新部103は、更新要求に基づいて、指定された構造化文書IDを指定し、構造化文書管理部200の排他制御部205に対して排他制御を要求する（ステップS400）。次に、インデックスリスト蓄積部13からインデックスリストを取得する（ステップS402）。さらに、総和管理部100の排他制御部305に対してインデックスリストに存在するインデックスに対して排他制御を要求する（ステップS404）。

【0115】

その後、構造化文書管理部200の更新部203に対して、対象となる構造化文書の全文の更新を要求する（ステップS406）。そして、処理の結果得られた差分情報をインデックス処理部300の更新部303に送り、インデックスを更新する旨要求する。なお、インデックスリストに存在する全てのインデックスに対して、インデックスを更新する旨要求する。

【0116】

次に、更新部303は、差分情報に基づいて、更新すべきインデックスを特定する（ステップS408）。すなわち、更新すべきインデックスの絶対経路IDを特定する。そして、更新部303は、インデックス蓄積部12に蓄積されているインデックス、すなわち絶対経路IDを更新する（ステップS410）。なお、本実施の形態においては、更新条件と、差分情報に基づいて、インデックスが更新される。

【0117】

ここで、差分情報とは、構造化文書の構造単位である。そして、差分情報は、当該ノードの絶対経路IDと共に示される。更新の対象が複数存在する場合は、複数の対象がリストとして表示される。

【0118】

所定のノードが完全に削除された場合、例えば、

DELETE :

document ("sample.xml") / AudioVisualSegment
[1] / TextAnnotation [2]

と表される。また、所定のノードの値が変更された場合には、例えば、

CHANGE :

document ("sample.xml") / AudioVisualSegment
[1] / TextAnnotation [3] / FreeTextAnnotation
[1] 修正後特許1

と表される。

【0119】

10

20

30

40

50

本発明におけるインデックスは、絶対経路IDとその値のみを有する簡単な形式であり、絶対経路IDも構造化文書の構造に基づいたものである。したがって、以下に示すように簡単にインデックスの更新処理を行うことができる。

【0120】

例えば、図13に示すように、兄弟ノードに存在するデータ「特許1」、「特許2」、「特許3」、「特許4」のうち、「特許2」の「FreeTextAnnotation」のノードが削除された場合には、「特許3」の「FreeTextAnnotation」の絶対経路IDの「FreeTextAnnotation[3]」の部分と、「FreeTextAnnotation[2]」に更新する。さらに、「特許4」の「FreeTextAnnotation」の絶対経路IDの「FreeTextAnnotation[4]」の部分と、「FreeTextAnnotation[3]」に更新する。このように、連続番号を変更するだけでよく、高速に更新処理を行うことができる。

10

【0121】

また、従来は、実際の構造化文書は、ノードが出現する位置、すなわち開始行目とインデント位置、終了行目とインデント位置の各情報を保持していた。そして、これらの情報により、問合せ結果のノードを取得した後、指定された問合せ結果ノードから相対経路で辿れるノードの値を取得していた。このため、構造化文書が変更された場合、構造化文書の構造とは無関係に、更新されたノードが出現する階層よりも下位の層の全てのインデックスを更新しなければならなかった。

【0122】

具体的には、例えば、要素「キーワード」における「キー」と「ワード」の間に改行を加えた場合には、XMLにおいて当該「キーワード」が記述された行以降のすべての内容を更新しなければならなかった。さらに、値ではなく、ノード単位の追加、削除に対してはこれに加えて、経路内IDの変更も必要であった。これに対して、本実施の形態では、差分情報と、絶対経路IDを更新するのみでよく、従来に比べて効率的に更新処理を行うことができる。

20

【0123】

なお、複数存在する場合であっても、リストの先頭から順に処理が行われた場合には、不整合が生じない。

【0124】

説明を図12に戻す。すべてのインデックスに対して、ステップS408およびステップS410の処理が完了すると（ステップS412, Yes）、インデックス処理部300の排他制御部305は、全インデックスに対する排他制御を解除し（ステップS414）、さらに、構造化文書管理部200の排他制御部205は、指定された構造化文書に対する排他制御を解除する（ステップS416）。そして、総合管理部100の通信部110は、更新処理が完了した旨をクライアントに通知する（ステップS418）。以上で、更新処理が終了する。

30

【0125】

次に、図14を参照しつつ、インデックス定義処理について説明する。インデックス定義処理とは、インデックスリスト蓄積部13に格納されているインデックスリストに新たにインデックスを定義する処理である。

40

【0126】

通信部110がクライアントからインデックスの内容を更新する旨の更新要求を受信し、要求振り分け部104に送る（ステップS500）。要求振り分け部104は、受け取った要求がインデックス定義である場合には（ステップS502, Yes）、更新部103が呼び出され、インデックス定義処理が開始する。具体的には、更新部103は、インデックスリスト蓄積部13に格納されているインデックスリストに対して、クライアントから指定されたインデックスを定義を追加する（ステップS504）。以上で、インデックス定義処理が終了する。

【0127】

50

次に、図15を参照しつつ、検索処理について説明する。検索処理とは、クライアントから指定された検索条件に合致するインデックスをクライアントに返す処理である。

【0128】

まず、通信部110がクライアントから検索要求を受信し、要求振り分け部104に送る（ステップS600）。要求振り分け部104は、受け取った要求が検索要求である場合には（ステップS601, Yes）、検索条件取得部105が呼び出され、検索処理を開始する。具体的には、検索条件取得部105は、検索要求に含まれる問い合わせ式を検索条件解析部106に送る。ここで、問い合わせ式とは、検索条件のことである。また、問い合わせ式は、インデックスの第1値と第2値のうち少なくとも一方に対応する条件が含まれている。

10

【0129】

次に、検索条件解析部106は、受け取った問い合わせ式を解析し（ステップS602）、解析された問合せ式、問合せキー、インデックスID（複数であればリスト）、要求された相対ノード式のリストを取得する（ステップS604）。

【0130】

ここで、相対ノード式とは、問い合わせ結果として特定されるノードのほかに、クライアントが取得したい情報を含むノードまでの相対経路である。すなわち、相対経路は、本発明における相対経路IDに相当する。

【0131】

例えば、「document("sample.xml")//FreeTextAnnotation[contains(text(),"特許2")]」で記述される問い合わせ式1に対して返された結果に対し、クライアントがさらに当該ノードの近辺の情報も結果として得たい場合がある。先の問合せ式1の結果であるノードを\$I1とすると、その結果得られたノードの近辺の情報を要求する問い合わせ式2は、「\$I1/./././././CreationInformation/Creation/Title」と記述することができる。このように、取得したいノードを、すでに特定したノードからの相対的な位置で表した経路情報が相対経路である。

20

【0132】

再び、説明を図15に示すフローチャートに戻す。検索条件解析部106の解析の結果、インデックスに基づいて、指定された検索対象を特定できる場合には、その解析された問合せ式、問合せキー、インデックスIDの組（複数であればリスト）を指定して（ステップS610）、インデックス処理部300の検索部304に検索を要求する（ステップS612）。この要求に対して、絶対経路IDとその値の組を取得する。なお、このとき、照合部315による文字列の照合により、特定される。

30

【0133】

一方、検索条件解析部106における解析の結果、指定された検索対象をインデックスを利用して特定することができない場合には（ステップS608, Yes）、問合せ式を指定し、構造化文書管理部200のノード値取得部204にノード値を要求する。これに対して、ノード値取得部204は、インデックス蓄積部12から既に登録済みの構造化文書のノード値を取得する。なお、ノード値は、絶対経路IDとその値の組である。また、該当するノードが複数ある場合には、そのリストである（ステップS620）。

40

【0134】

次に、ノード値取得部204は、問い合わせ式に該当するノードの絶対経路IDと値のリストを取得する（ステップS622）。なお、全構造化文書から、問い合わせ式に該当する全てのノードを取得するまで繰り返し（ステップS624）、ステップS630に進む。

【0135】

次に、相対ノードパス数が0であれば（ステップS630, No）、絶対経路IDとその値の組（複数であればリスト）をクライアントに返す（ステップS632）。相対ノードパス数が0でなければ（ステップS630, Yes）、相対ノード式リストを取得し（ス

50

テップ S 6 4 0)、相対ノード式リストとともに、絶対経路 I D とその値の組(複数であればリスト)をクライアントに返す(ステップ S 6 3 2)。以上で、検索処理が終了する。

【0136】

問い合わせ式に対する結果は、所定のノードを示すリストである。しかし、問合せ結果として、そのノードから相対経路によって辿れるノードが必要とされる場合もある。

【0137】

例えば、図 5 に示した構造化文書 5 1 0 に対して「“Movie”」を含む文字列を取得するという問い合わせ式、すなわち図 1 6 (A) に示す第 1 問い合わせ式 5 4 0 に対して、文字列「“Movie Tool サンプルストリーム”」、「“オープニング Movie”」、「“Movie Tool の動作”」およびが結果として返される。

10

【0138】

この結果には、構造化文書における構造に関する情報が含まれていないので、例えば、この問い合わせの後、さらに「“Movie Tool サンプルストリーム”」の上位の階層の情報を取得したい場合に、再度同様の問い合わせを行わなければならない。すなわち、再度図 1 6 (B) の第 2 問い合わせ式 5 4 2 を指定して同様の処理を行う必要があり、手続きが面倒であった。

【0139】

これに対して、図 1 5 を参照しつつ説明した本実施の形態における検索処理によれば、文書管理装置 1 0 は、図 1 6 (A) に示される第 1 問い合わせ式 5 4 0 に対して、文字列「“Movie Tool サンプルストリーム”」と「“オープニング Movie”」が結果として返されるとともに、これらの絶対経路 I D が返される。また、この相対経路を自動的に絶対経路 I D に変換するので、利用者は、第 2 の問い合わせ式を改めて指定する必要はなく、手続きを簡略化することができる。

20

【0140】

以上のように、インデックスの第 1 値および第 2 値の少なくとも一方に対する条件が指定された検索条件に対して、該当する値がリスト構造内に含まれている場合には、このインデックスを全て検索結果として返すことができる。

【0141】

また、検索条件として、複数の検索条件と、各検索条件に対する論理演算子とを取得した場合には、まず各検索条件とインデックス蓄積部 1 2 に管理されているインデックスとをそれぞれ照合する。そして、照合結果に基づいて、各検索条件に合致する前記インデックスをそれぞれ特定する。さらに、各検索条件に対して特定した複数のインデックスに対し、論理演算子を用いた論理演算を行い、当該論理演算の結果得られたインデックスを、検索条件に合致する前記インデックスとして特定する。

30

【0142】

さらに、検索条件として取得したインデックスは文字列で構成されているので、各検索条件の文字列とインデックス蓄積部 1 2 に蓄積されているインデックスに含まれる文字列とを照合し、文字列の照合の結果に基づいて、各検索条件に合致する前記インデックスをそれぞれ特定する。そして、特定した複数のインデックス間で文字列を照合することにより論理演算を行う。そして、当該論理演算の結果得られたインデックスを、検索条件に合致するインデックスとして特定する。

40

【0143】

図 1 7 は、図 1 5 において説明した相対ノード値取得ステップ(ステップ S 6 4 0)における詳細な処理を示すフローチャートである。

【0144】

総管理部 1 0 0 の検索条件取得部 1 0 5 は、既に取得している絶対経路 I D と相対ノード式リストのうちから(ステップ S 7 0 0)、最初に処理すべきの絶対経路 I D を取得する(ステップ S 7 0 2)。さらに、最初に処理する相対ノード式を取得する(ステップ S 7 0 4)。そして、取得した絶対経路 I D と相対ノード式に基づいて、絶対経路で示さ

50

れた問い合わせ式を生成する（ステップS706）。

【0145】

図18（A）は、検索条件取得部105が取得した絶対経路IDリストの一例を示している。また、図18（B）は、検索条件取得部105が取得した相対ノード式リストの一例を示している。図18（A）に示す絶対経路ID600のそれぞれと図18（B）に示す相対ノード式リストとの組み合わせからそれぞれ問い合わせ式が生成される。

【0146】

図19は、図18に示す絶対経路IDと相対ノード式リストから生成された問い合わせ式リストを示している。図19（A）に示す問い合わせ式リスト620のうち上から1行目から3行目の問い合わせ式622は、絶対経路IDリスト600の1行目の絶対経路IDと、相対ノード式リスト610とにより生成された問い合わせ式である。また、4行目から6行目の問い合わせ式624は、絶対経路IDリスト600の2行目の絶対経路IDと、相対ノード式リスト610とにより生成された問い合わせ式である。

【0147】

また、図19（B）は、図19（A）に示す問い合わせ式を簡略化した簡略問い合わせ式リスト630を示している。このように、問い合わせ式の簡略化して表現してもよい。

【0148】

再び、説明を図16に戻す。問い合わせ式が生成されると（ステップS706）、構造化文書管理部200のノード値取得部204は、生成された問い合わせ式に基づいて、ノード値を取得する（ステップS708）。総管理部100の検索条件取得部105は、現在対象としている絶対経路IDに対し、未処理の相対ノード式がある場合には（ステップS710、No）、次の相対ノード式を取得し（ステップS712）、再び問い合わせ式を生成する（ステップS706）。以上、全ての相対ノード式について生成した問い合わせ式に対するノード値を取得するまで、ステップS706およびステップS710の処理を繰り返す（ステップS710）。

【0149】

全ての相対ノード式について処理が完了すると（ステップS710、Yes）、次の絶対経路IDに対して、同様の処理を行う。すなわち、未処理の絶対経路IDがある場合には（ステップS720、No）、次の絶対経路IDを取得し（ステップS722）、取得した絶対経路IDに対して組み合わせる一の相対ノード式を再び取得する（ステップS704）。以上、全ての絶対経路IDについて生成した問い合わせ式に対するノード値を取得するまで、ステップS704からステップS720の処理を繰り返す。以上で、相対ノード値取得処理が終了する。

【0150】

このように、ノード値取得部204は、指定された問合せ式に基づいて、一つの構造化文書から、該当するノードの絶対経路IDとその値を取得することができる。

【0151】

一般に問合せの対象となる構造化文書の数が増えると、すなわち、構造化文書蓄積部11に蓄積されている構造化文書の数が増えると、構造化文書の数増加に伴って、ファイルI/OやDOMツリー生成の問題で処理時間が長くなる。

【0152】

これに対して、本実施の形態においては、図15から図17を参照しつつ説明したように、インデックスを利用できる場合は、必ずインデックスを利用して問合せを行うので、処理時間を短縮することができる。

【0153】

クライアントが問合せの結果の件数を得て、その問合せ結果の詳細な情報を得たい場合、その件数は、10件程度であることが多い。10件程度であれば、仮に全て別の構造化文書であったとしても、全構造化文書から値を取得したり、多数のインデックスからノード値を取得するよりも、相対ノード式によってノード値を取得するほうが高速に処理が行えるためである。

10

20

30

40

50

【0154】

次に、図15において説明した問い合わせ式解析処理（ステップS602）について、図20を参照しつつ詳細に説明する。図20は、問い合わせ式の一例を示している。

【0155】

総管理部100の検索条件解析部106は、まず、問い合わせ式をインデックスを利用できる可能性のある単位に分解する。具体的には、以下の順序で処理を行う。

Step 1 問い合わせ式本体と相対ノード式とを分解する

Step 2 問合せ式本体をAxis（軸）ごとに分解する

Step 3 各Axis毎の属性情報を取得する。

ここで、属性情報とは、要素名、軸方向と数、述語種（比較演算子）とその値のことである。属性情報のうち要素名とは、「Who」、「Name」などであり、「. . .」や「ancestor-or-self::」は空のノードと見なされる。軸方向とは、そのパスが構造のどちらに向いているかを示すもので、要素名だけであれば、軸方向は“下り”で、その数1であることを示している。

【0156】

また、「. . .」であれば軸方向“上り”その数1となり、ancestor-or-self::AudioVisualSegment[1]であれば軸方向“上り”その数1となる。述語種とは、「contains(text(),)」などその問合せ条件に該当し、その問合せキーは「東京」となる。述語に該当するものがない場合は空となる。

【0157】

まず、図20に示す問い合わせ式の問い合わせ式本体「document("sample.xml")//Who/Name[contains(text(),"山田")]//. . .//Where/Name[contains(text(),"東京")]//ancestor-or-self::AudioVisualSegment[1]」と残りの部分、すなわち相対ノード式とを分離する。さらに、問い合わせ式本文は、「/」文字で区切られた軸毎に分解され、その属性情報が取得される。分解されたAxisは以下の条件によって、まとめられる。解析の結果まとめられたものを軸集合と呼ぶ。

Condition 1 軸方向が上りの軸と、軸方向が下りの軸を分岐点としてまとめる。

Condition 2 軸方向が下りで述語種が空ではない軸と、軸方向が下りの軸を分岐点としてわけける。

【0158】

各軸集合の中で以下の条件を満たすものが、インデックスを利用できる軸の候補となる。

1 述語種が空でないもの

2 次の軸の軸方向が上りである軸方向が下りの軸

インデックスを利用できる候補の軸の要素名を条件にし、インデックスリストにインデックスIDを取得する。その際、前の軸が下りでその数が1の場合は、その要素名も連結し、そのインデックスが存在すればそちらを採用する。例えば、NameよりもWho/Nameのインデックスを優先する。

【0159】

インデックスIDが存在しない場合、その要素名に対するインデックスは存在しないことになり、利用可能なインデックスが存在しないという結果を返す。存在すれば、そのインデックスID、解析した問合せ式、相対ノード式を返す。

【0160】

続いて、インデックス処理部300の検索部304は、解析された問合せ式（軸集合）、問合せキー、インデックスIDの組（複数であればリスト）の指定に基づいて、各組毎に、インデックス蓄積部12対して問合せキーを第1値の条件として取得を要求する。そして、該当するものがあればその値と絶対経路IDの組のリストを取得する。これを各組毎に繰り返す。

【0161】

10

20

30

40

50

ここで、値と絶対経路IDの組のリストを取得する方法について詳述する。例えば、問い合わせ式が、論理積演算である場合には、そのリストの絶対経路IDが重複するもののみ結果として取得する。また、論理和演算であれそのリストの絶対経路IDが重複するものを一つだけ結果として取得する。

【0162】

図3に示した構造化文書に対して、図21に示す問い合わせ式が指定された場合の処理を例に説明する。

【0163】

図21に示すように、2つの条件の論理積である場合、まず、各条件に合致するノードの絶対経路IDを取得する。図21に示す2つの条件に対しては、結果1の絶対経路ID「AudioVisualSegment[1] / Creation Information[1] / Creation[1] / Title[1]」、結果2の絶対経路ID「AudioVisualSegment[2] / Creation Information[1] / Creation[1] / Title[1]」、および結果3の絶対経路ID「AudioVisualSegment[1] / TemporalDecomposition[1] / AudioVisualSegment[1] / TemporalDecomposition[1] / AudioVisualSegment[1] / TextAnnotation[1] / FreeTextAnnotation[1]」を取得する。

【0164】

以上、結果1から結果3の論理積を求め、問合せ式を満たす「AudioVisualSegment」を取得するためには、

【0165】

▲1▼ 結果1の絶対経路IDの先頭側の文字列「AudioVisualSegment[1]」と結果3の絶対経路IDの先頭側文字列「AudioVisualSegment[1]」までを比較し、等しいので問い合わせ式を満たす。

▲2▼ 結果1の絶対経路IDの先頭側の文字列「AudioVisualSegment[2]」と結果3の絶対経路IDの先頭側文字列「AudioVisualSegment[1]」までを比較し、等しいので問い合わせ式を満たす。

以上、2回の文字列処理、すなわち、文字列の照合により問い合わせ式に指定された論理演算を行うことができる。

【0166】

なお、従来は、問合せ式に該当する2つの「Title」ノードと1つの「FreeTextAnnotation」ノードを取得することはできるが、両方の条件を満たした「AudioVisualSegment」要素を取得するためには、まず「Title」ノードを取得し、ポインタで「AudioVisualSegment」まで辿り、更に総当りで計五つの「FreeTextAnnotation」が条件を満たすか問合せる必要がある。

【0167】

以上のように、経路が異なる複数の経路情報も含めた問合せ式が複合されている論理積演算の処理に対し、本実施の形態においては、インデックスに含まれた絶対経路IDの文字列処理のみで行うことができる。従って、従来のようにポインタを辿る処理に比べて処理を効率化することができる。

【0168】

図22は、文書管理装置10のハードウェア構成を示す図である。文書管理装置10は、ハードウェア構成として、文書管理装置10における文書管理の処理を実行するプログラムなどが格納されているROM52、ROM52内のプログラムに従って文書管理装置10の各部を制御し、文書管理等を実行するCPU51、ワークエリアが形成され、文書管理装置10の制御に必要な種々のデータが記憶されているRAM53、ネットワークに接続して、通信を行う通信I/F57、および各部を接続するバス62を備えている。

【0169】

先に述べた文書管理装置10における文書管理処理を実行する文書管理プログラムは、インストール可能な形式又は実行可能な形式のファイルでCD-ROM、フロッピー(R)ディスク(FD)、DVD等のコンピュータで読み取り可能な記録媒体に記録されて提供される。

【0170】

また、本実施形態の文書管理プログラムを、インターネット等のネットワークに接続されたコンピュータ上に格納し、ネットワーク経由でダウンロードさせることにより提供するように構成しても良い。

【0171】

この場合には、文書管理プログラムは、文書管理装置10において上記記録媒体から読み出して実行することにより主記憶装置上にロードされ、上記ソフトウェア構成で説明した各部が主記憶装置上に生成されるようになっている。

【0172】

(実施の形態2)

次に、実施の形態2にかかる文書管理装置10について説明する。図23は、実施の形態2にかかる文書管理装置10の構成を示すブロック図である。

【0173】

実施の形態2にかかる文書管理装置10は、実施の形態1にかかる文書管理装置10の構造化文書蓄積部11にかえて構造化文書ファイルパス蓄積部14を備える。また、構造化文書管理部200の排他制御部205、インデックス処理部300の排他制御部305を備えない。また、総合管理部100は、インデックスリスト12に蓄積されているインデックスリストを再構築するインデックス再構築部111を備える。

【0174】

構造化文書ファイルパス蓄積部14は、構造化文書にかえて、構造化文書へのファイルパスを蓄積している。構造化文書ファイルパス蓄積部14は、構造化文書の追加要求に対して、構造化文書ファイルパスを蓄積する。なお、ノード値取得部204からの取得要求に対して、構造化文書を識別するための構造化文書IDは、各構造化文書ファイルパスである。構造化文書ファイルパスは、ファイルとして蓄積されてもよく、また他の例としては、メモリ上に蓄積されてもよい。

【0175】

構造化文書ファイルパス蓄積部14は、構造化文書IDを指定された取得要求に対し、該当する構造化文書ファイルパスが存在すれば、該当する構造化文書ファイルパスを返す。また、構造化文書IDを指定された削除要求に対し、該当する構造化文書ファイルパスが存在すれば、その構造化文書ファイルパスを削除し、管理対象から除く。本実施の形態における構造化文書ファイルパス蓄積部14は、計算機のファイルシステムで構成されている。なお、他の例としては、構造化文書ファイルパス蓄積部14は、データベースであってもよい。

【0176】

なお、構造化文書管理部200の登録部201、削除部202、および更新部203は、構造化文書ファイルパス蓄積部14に蓄積されている構造化文書ファイルパスに対してそれぞれ所定の処理を行う。

【0177】

実施の形態2においては、総合管理部100の要求振り分け部104が受け付ける要求種にインデックス再構築が加えられる。要求振り分け部104がインデックス再構築の要求を取得した場合、総合管理部100のインデックス再構築部111が呼び出される。

【0178】

このときの基本的な処理は、図8から図11を参照しつつ説明した実施の形態1にかかる文書管理装置10の処理と同様であるが、実施の形態2においては、排他制御は行われない。

10

20

30

40

50

【0179】

なお、実施の形態2においては、構造化文書IDは、構造化文書が参照できるファイルパスに固定される。また、構造化文書への操作は、そのファイルパスを参照して行う。

【0180】

総管理部100のインデックス再構築部111は、要求振り分け部104から要求を受け、削除部302に対して全ノードのインデックス削除を要求する。その後、ノード値取得部204に対して、全構造化文書のファイルパス取得の要求を行い全構造化文書のファイルパスを取得し、インデックスリスト蓄積部13からインデックスリストを取得する。その後、インデックス処理部300の追加部301は、インデックスリストに定義されている数の分のインデックスを追加する。なお、この処理を全構造化文書に対して行う。

10

【0181】

なお、これ以外の文書管理装置10の構成および動作は、図1から図20を参照しつつ説明した実施の形態1にかかる文書管理装置10の構成および動作と同様である。

【0182】

本実施の形態においては、全インデックスを再構築しているが、構造化文書ファイルパス蓄積部14において、各ファイルパスに対応して、最終更新日時とファイルサイズ、UNIX(R)ファイルシステムという1ノードなど、その構造化文書のファイルが更新されたことを確認できる情報も保持し、インデックス再構築要求時に更新された構造化文書のみインデックスを削除、再構築するということも容易に考えられる。

【0183】

(実施の形態3)

次に、実施の形態3にかかる文書管理装置10について説明する。図24は、実施の形態3にかかる文書管理装置10の構成を示すブロック図である。

20

【0184】

実施の形態3にかかる文書管理装置10は、実施の形態1において説明したインデックス処理部300とインデックス蓄積部12とを複数備えている。すなわち、実施の形態3にかかる文書管理装置10は、インデックスを蓄積する第1インデックス蓄積部121、第2インデックス蓄積部122、および第3インデックス蓄積部123と、各蓄積部に蓄積されているインデックスに対し、それぞれ処理を行う、第1インデックス処理部300、第2インデックス処理部310、および第3インデックス処理部320とを備えている。

30

【0185】

さらに、実施の形態3においては、第1インデックス処理部300および第1インデックス蓄積部121は、文字列専用である。また、第2インデックス処理部310および第2インデックス蓄積部122は、数値専用である。また、第3インデックス処理部320および第3インデックス蓄積部123は、画像特徴量専用である。

【0186】

このように、複数のインデックス処理部およびインデックス蓄積部を有しているので、データの種別に応じて、複数のインデックス処理部およびインデックス蓄積部を使い分けることができる。

【0187】

なお、どの場合にどのインデックス処理部を利用するかは、インデックスリスト蓄積部が保持しているインデックスIDにより決定される。

40

【0188】

これ以外の文書管理装置10の構成および動作は、図1から図22を参照しつつ説明した実施の形態1にかかる文書管理装置10の構成および動作と同様である。

【0189】

また、実施の形態3においては、複数のインデックス蓄積部それぞれに対応してインデックス処理部が設けられていたが、他の例としては、複数のインデックス蓄積部に対して一のインデックス処理部が設けられていてもよい。この場合にも、処理を効率化することができる。

50

【0190】

以上、本発明を実施の形態を用いて説明したが、上記実施の形態に多様な変更または改良を加えることができる。

【0191】

【発明の効果】

以上説明したように、請求項1にかかる発明によれば、特定した要素が構造化文書内のいずれの位置に格納されているのかを把握することができるので、例えば、さらに他の要素を特定する場合に、既に特定している要素の位置を基準とした、他の要素までの相対経路に基づいて他の要素を特定することができる。このように、特定した要素の位置に基づいて、検索処理を行うことにより、処理の効率化を図ることができるという効果を奏する。

10

【0192】

また、請求項2にかかる発明によれば、インデックス管理手段は、要素とともにインデックスを特定することができる。また、このインデックスは、構造化文書における要素の位置を示しているもので、これにより、要素の位置を把握することができる。このように、特定した要素の位置に基づいて、処理を行うことにより、処理の効率化を図ることができるという効果を奏する。

【0193】

また、請求項3にかかる発明によれば、インデックスは要素の位置を示しているもので、要素の位置を示す検索条件を取得した場合には、検索条件とインデックスとの照合結果に基づいて、検索条件に合致する要素を特定することができる。従って、より効率的に検索対象を特定することができるという効果を奏する。

20

【0194】

また、請求項4にかかる発明によれば、同一の経路で特定される要素が複数ある場合、すなわち一の親ノードに対して、複数の子ノードが存在するような構成であっても、複数の要素それぞれを識別することができる。従って、すべての要素を正確に識別することができるという効果を奏する。

【0195】

また、請求項5にかかる発明によれば、検索条件が要素位置特定経路情報および経路内要素特定情報のうち少なくとも一方を含む場合には、当該検索条件とインデックスとを照合することによって、検索条件に合致する要素を特定することができる。従って、より効率的に検索対象を特定することができるという効果を奏する。

30

【0196】

また、請求項6にかかる発明によれば、インデックスと検索対象との照合により検索対象を特定することができるので、処理効率を向上させることができるという効果を奏する。

【0197】

また、請求項7にかかる発明によれば、文字列照合の結果を利用することにより、検索処理の効率を向上させることができるという効果を奏する。

【0198】

また、請求項8にかかる発明によれば、絶対経路特定手段は、前回特定したインデックスと、検索条件取得手段が取得した相対経路に基づいて、次に特定すべき要素の絶対経路、すなわちインデックスを特定することができるので、検索条件を解析して構造化文書の根から検索を開始する場合に比べて、検索処理の効率を向上させることができるという効果を奏する。

40

【0199】

また、請求項9にかかる発明によれば、この請求項9の発明によれば、絶対経路特定手段によって特定された絶対経路に基づいて、検索条件に合致する要素を特定することができる。すなわち、絶対経路を取得しなくとも、相対経路から特定した絶対経路に基づいて、要素を特定することができるという効果を奏する。

【0200】

また、請求項10にかかる発明によれば、前回特定したインデックスと今回特定したイン

50

デックスとの排他論理和を算出することにより、相対経路を特定することができる。すなわち、比較的簡単な処理により相対経路を特定することができるという効果を奏する。

【0201】

また、請求項11にかかる発明によれば、検索条件に合致するとして抽出した要素と、当該要素のインデックスとを対応付けて出力するので、ユーザは、要素が構造化文書のいずれの位置に格納されているのかを把握することができるという効果を奏する。

【0202】

また、請求項12にかかる発明によれば、インデックス蓄積手段に蓄積されているインデックスを更新する更新手段をさらに備えているので、インデックスを自由に更新することができるという効果を奏する。

【0203】

また、請求項13にかかる発明によれば、インデックス蓄積手段に蓄積されているインデックスの登録および削除を行う登録削除手段をさらに備えているので、インデックスを自由に登録および削除することができるという効果を奏する。

【0204】

また、請求項14にかかる発明によれば、更新手段が一の処理を行っているときに、排他制御手段は、他の処理を排除するので、複数の処理が同時に行われることにより、不適切な更新が行われるのを避けることができるという効果を奏する。

【0205】

また、請求項15にかかる発明によれば、一の処理により、他のデータとの整合性がとれなくなるのを回避することができるという効果を奏する。

【0206】

また、請求項16にかかる発明によれば、インデックスは要素の位置を示しているので、要素の位置を示す検索条件を取得した場合には、検索条件とインデックスとの照合結果に基づいて、検索条件に合致する要素を特定することができる。従って、より効率的に検索対象を特定することができるという効果を奏する。

【0207】

また、請求項17にかかる発明によれば、複数のインデックス蓄積手段を備えているので、処理の効率化を図ることができるという効果を奏する。

【0208】

また、請求項18にかかる発明によれば、複数のインデックス蓄積手段と、各インデックス蓄積手段を管理する複数のインデックス管理手段とを備えているので、分散処理が可能となり、処理の効率化を図ることができるという効果を奏する。

【0209】

また、請求項19にかかる発明によれば、インデックスは要素の位置を示しているので、要素の位置を示す検索条件を取得した場合には、検索条件とインデックスとの照合結果に基づいて、検索条件に合致する要素を特定することができる。従って、より効率的に検索対象を特定することができるという効果を奏する。

【0210】

また、請求項20にかかる発明によれば、特定した要素が構造化文書内のいずれの位置に格納されているのかを把握することができるので、例えば、さらに他の要素を特定する場合に、既に特定している要素の位置を基準とした、他の要素までの相対経路に基づいて他の要素を特定することができる。このように、特定した要素の位置に基づいて、検索処理を行うことにより、処理の効率化を図ることができるという効果を奏する。

【0211】

また、請求項20にかかる発明によれば、特定した要素が構造化文書内のいずれの位置に格納されているのかを把握することができるので、例えば、さらに他の要素を特定する場合に、既に特定している要素の位置を基準とした、他の要素までの相対経路に基づいて他の要素を特定することができる。このように、特定した要素の位置に基づいて、検索処理を行うことにより、処理の効率化を図ることができるという効果を奏する。

10

20

30

40

50

【図面の簡単な説明】

【図 1】実施の形態 1 である文書管理装置 10 の構成を示すブロック図である。

【図 2】構造化文書蓄積部 11 に蓄積されている構造化文書のデータ構成を模式的に示す図である。

【図 3】MPEG-7 で記述した構造化文書を示す図である。

【図 4】動画データにおける一部のデータを指定する記述を示す図である。

【図 5】実施の形態 1 における動画データの構造化文書 510 を示す図である。

【図 6】XPath 式により示された絶対経路 ID を示す図である。

【図 7】インデックス蓄積部 12 に格納されているインデックスのデータ構造を示す図である。

10

【図 8】文書管理装置 10 の新規文書登録処理を示すフローチャートである。

【図 9】図 8 におけるインデックス追加処理（ステップ S110）における詳細な処理を示すフローチャートである。

【図 10】構造化文書を示す図である。

【図 11】文書管理装置 10 の削除処理を示すフローチャートである。

【図 12】文書管理装置 10 の更新処理を示すフローチャートである。

【図 13】インデックスの削除を説明するための図である。

【図 14】文書管理装置 10 のインデックス定義処理を示すフローチャートである。

【図 15】文書管理装置 10 の検索処理を示すフローチャートである。

【図 16】図 15 において説明した相対ノード値取得ステップ（ステップ S640）における詳細な処理を示すフローチャートである。

20

【図 17】問い合わせ式を示す図である。

【図 18】絶対経路 ID リストの一例を示す図である。

【図 19】図 18 に示す絶対経路 ID と相対ノード式リストから生成された問い合わせ式リストを示す図である。

【図 20】問い合わせ式の一例を示す図である。

【図 21】問い合わせ式の一例を示す図である。

【図 22】文書管理装置 10 のハードウェア構成を示す図である。

【図 23】実施の形態 2 に係る文書管理装置 10 の構成を示すブロック図である。

【図 24】実施の形態 3 に係る文書管理装置 10 の構成を示すブロック図である。

30

【符号の説明】

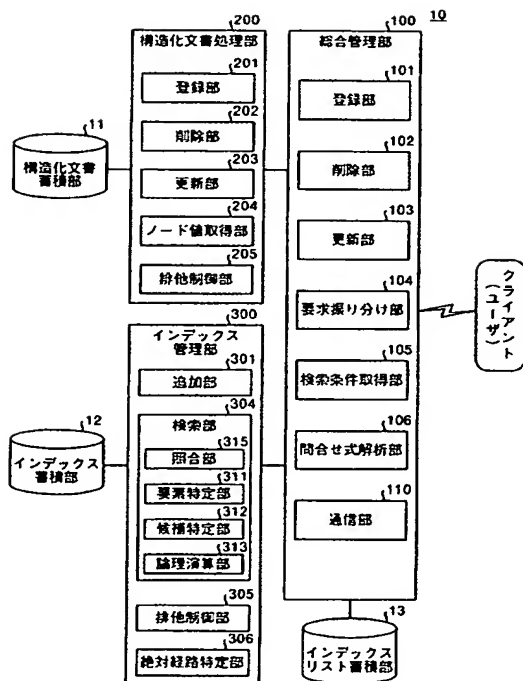
- 10 文書管理装置
- 11 構造化文書蓄積部
- 12 インデックス蓄積部
- 13 インデックスリスト蓄積部
- 14 構造化文書ファイルパス蓄積部
- 100 総合管理部
- 101 登録部
- 102 削除部
- 103 更新部
- 105 検索条件取得部
- 106 検索条件解析部
- 110 通信部
- 111 インデックス再構築部
- 121 第 1 インデックス蓄積部
- 122 第 2 インデックス蓄積部
- 123 第 3 インデックス蓄積部
- 200 構造化文書管理部
- 201 登録部
- 202 削除部

40

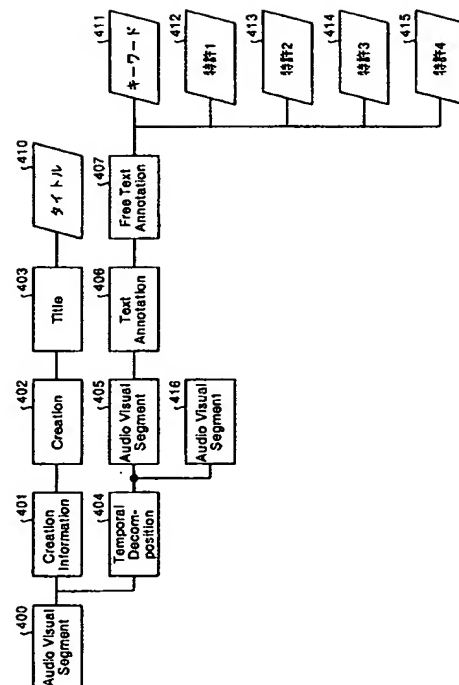
50

- 203 更新部
- 204 ノード値取得部
- 205 排他制御部
- 300 インデックス処理部
- 301 追加部
- 302 削除部
- 303 更新部
- 304 検索部
- 305 排他制御部
- 306 相対経路特定部
- 315 照合部
- 311 要素特定部
- 312 候補特定部
- 314 論理演算部

【図1】



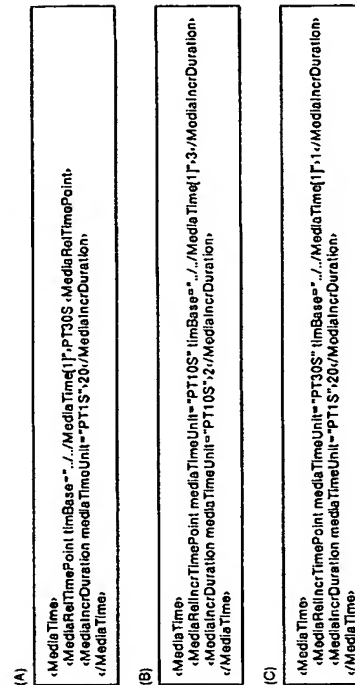
【図2】



【 図 3 】



【 図 4 】



【 図 5 】



【 図 6 】

(A) 文書名 sample.xml

```

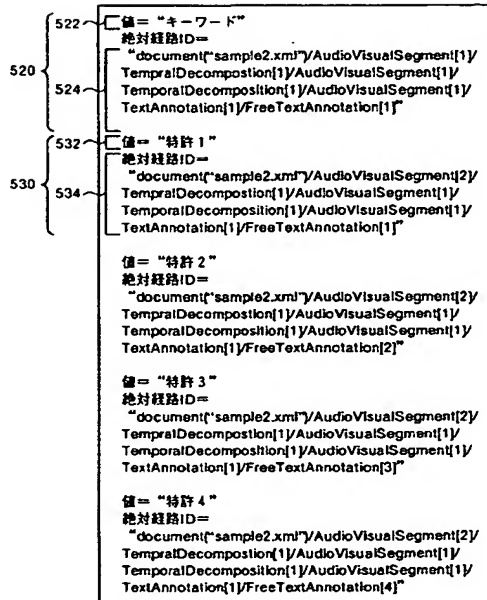
<?xml>
  <StructuralUnit>
    <Name>Pattern</Name>
  </StructuralUnit>
  <Location>
    <Name>東京</Name>
  </Location>
  <Location>
    <Name>大阪</Name>
  </Location>
  
```

(B)

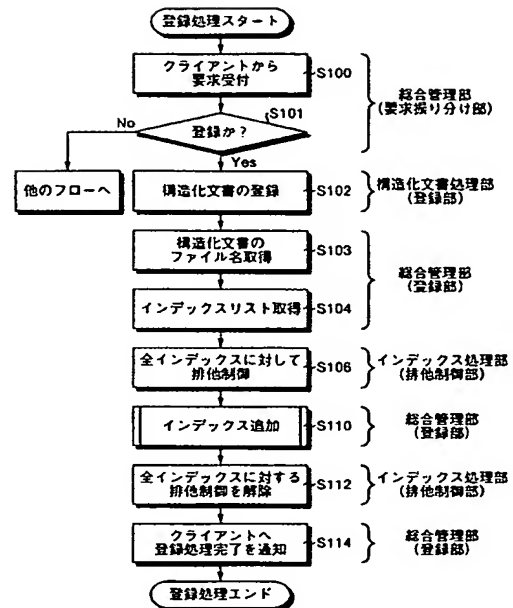
```

document("sample.xml")/StructuralUnit[1]
document("sample.xml")/StructuralUnit[1]/Name[1]
document("sample.xml")/Location[1]
document("sample.xml")/Location[1]/Name[1]
document("sample.xml")/Location[2]
document("sample.xml")/Location[2]/Name[2]
  
```

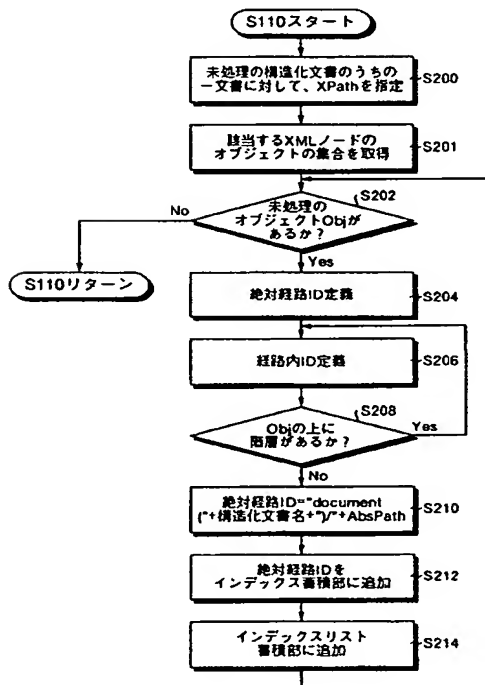
【図 7】



【図 8】



【図 9】



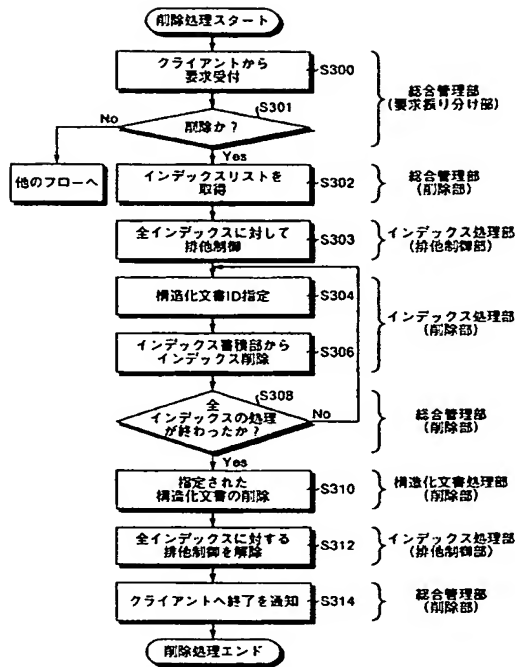
【図 10】

```

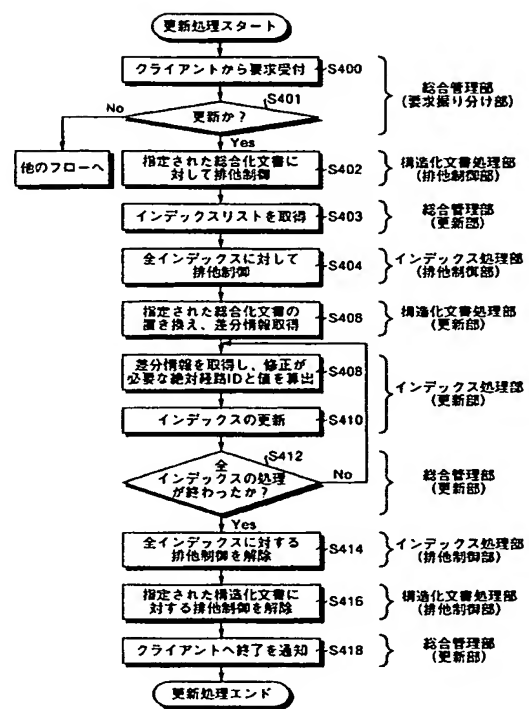
<StructuralUnit>
  <Name>Pattern</Name>
</StructuralUnit>
<Location>
  <Name>東京</Name>
</Location>

```

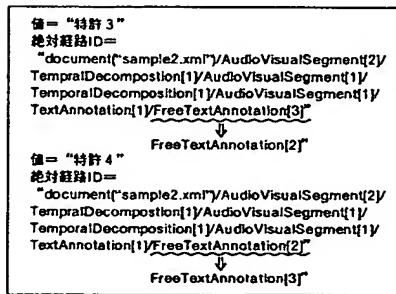

【図11】



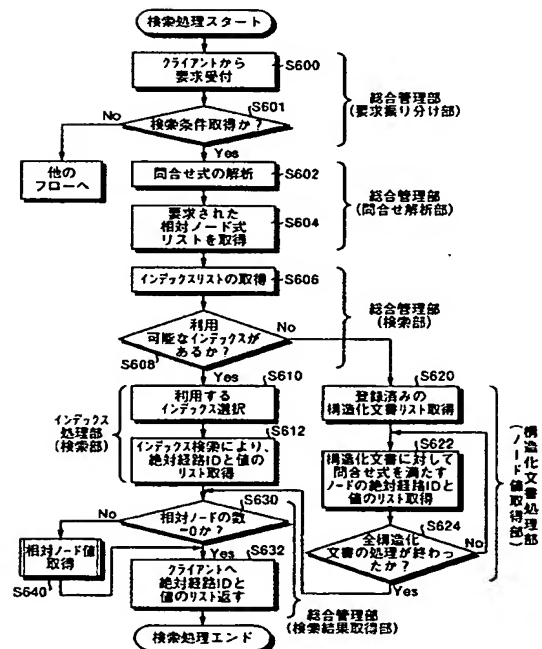
【図12】



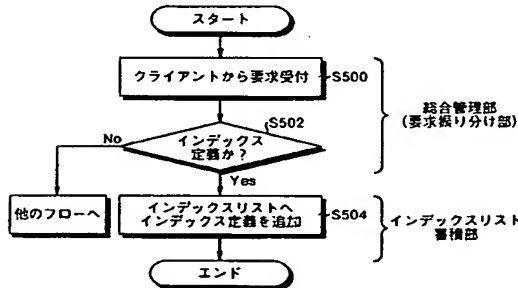
【図13】



【図15】



【図14】

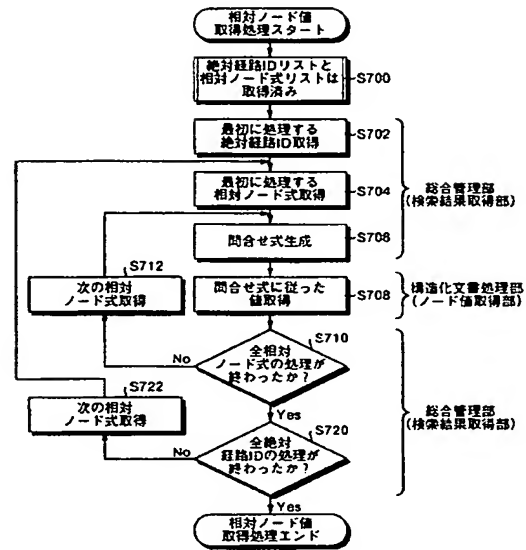


【図 16】

(A) /Mpeg7//TextAnnotation/FreeTextAnnotation[contains(text(),"Movie")] 540

(B) /Mpeg7//TextAnnotation/FreeTextAnnotation[contains(text(),"Movie")]
/../../MediaTime/MediaRelTimePoint[rich:timeCompare("=10s")] 542

【図 17】

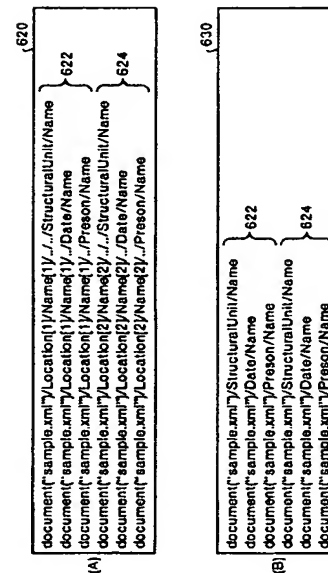


【図 18】

(A) document(["sample.xml"/Location[1]/Name[1]
document(["sample.xml"/Location[2]/Name[2]) 600

(B) \$1/../../StructuralUnit/Name
\$1/../../Date/Name
\$1/../../Preson/Name 610

【図 19】



【図20】

```

for $i in document("sample.xml")//Who/Name[contains(text(),"山田")]/../
Where/Name[contains(text(),"東京")]/ancestor-or-self::
AudioVisualSegment{1}
return
$!//TextAnnotation/StructuredAnnotation/Who/Name
$!//TextAnnotation/StructuredAnnotation/Where/Name
$!//TextAnnotation/StructuredAnnotation/When/Name

```

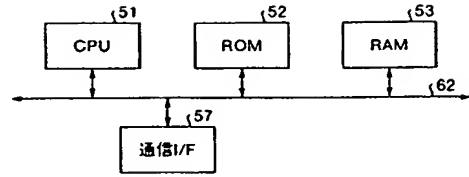
【図21】

```

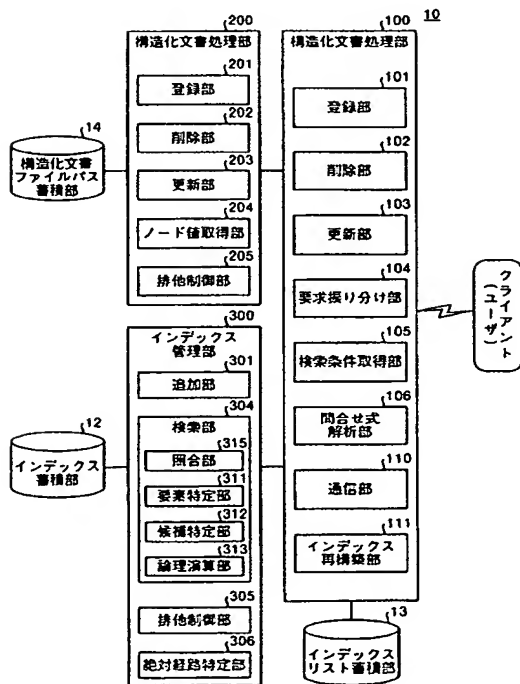
/AudioVisualSegment/([CreationInformation/Creation[Title="タイトル"]
and
(TemporalDecomposition//FreeTextAnnotation[contains(Text(),"キー")])

```

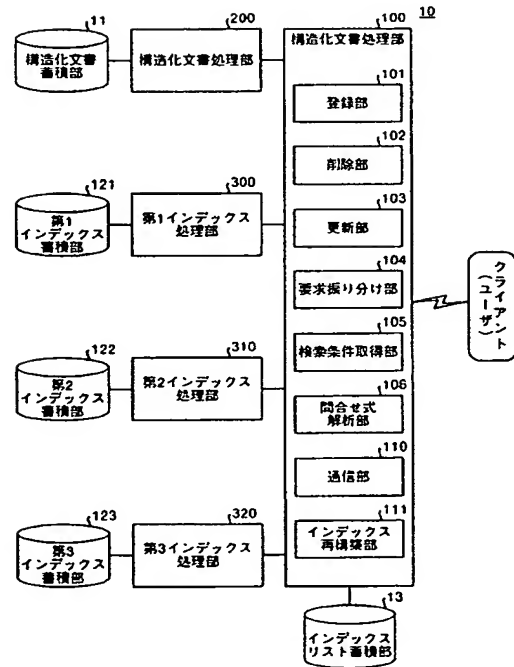
【図22】



【図23】



【図24】



フロントページの続き

(72)発明者 山形 純一

東京都大田区中馬込 1 丁目 3 番 6 号 株式会社リコー内

(72)発明者 小山 毅

東京都大田区中馬込 1 丁目 3 番 6 号 株式会社リコー内

Fターム(参考) 5B009 QA09 SA03 VA02 VA09

5B075 ND03 ND20 ND35 NK02 NK46 UU06

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record.

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.